Identification of the muscoride biosynthetic pathway leads to the discovery of a novel antimicrobial natural product

Rose-Marie Andsten

Master's thesis

University of Helsinki

Department of Microbiology

Microbiology

June 2020

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty	Laitos/Institution – Department	
Faculty of Agriculture and Forestry	Department of Microbiology	
Tekijä/Författare – Author		
Rose-Marie Andsten		
Työn nimi/Arbetets titel – Title		
Identification of the muscoride biosynthetic pathway leads to the discovery of a novel antimicrobial natural product		
Oppiaine/Läroämne – Subject		
Microbiology		
Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/Sidoantal – Number of pages
Master's Thesis	June 2020	60 + 8 appendices

Tiivistelmä/Referat – Abstract

Bacteria are a great source of natural products with complex chemical structures and diverse biological activities. Many have therapeutic properties and half of drugs in clinical use today are derived directly or indirectly from natural products. The pharmaceutical industry stopped investing in drug development from natural resources, due to perceived limitations in chemical space, and difficulties in rediscovery of known compounds and in obtaining sufficient quantities of natural products for clinical trials. There is now renewed interest in natural products as drug leads driven by technological advances in genome sequencing and analytical chemistry. Cyanobacteria produce a variety of natural products with therapeutic potential. Muscoride A is an unusual peptide alkaloid produced by a terrestrial freshwater cyanobacterium with reported antimicrobial activity.

The aim of this study was to characterize the biosynthetic origin and biological activity of muscoride A. I identified the 12.7 kb muscoride (*mus*) biosynthetic gene cluster from a draft genome of *Nostoc* sp. PCC 7906 using bioinformatics analysis. The *mus* biosynthetic gene cluster encoded enzymes for the heterocyclization, oxidation and prenylation of a precursor protein. Comparative genomics identified a *mus* biosynthetic gene cluster in the unpublished draft genome of *Nostoc* sp. UHCC sp. 0398 encoding a novel muscoride. This novel muscoride, muscoride B, was detected from *Nostoc* sp. UHCC 0398 based on this analysis. Muscoride B was purified using solid phase extraction and high-performance liquid chromatography and the chemical structure was verified by combining nuclear magnetic resonance and mass spectrometry data. Furthermore, the function and evolutionary history of the muscoride prenyltransferases were studied. A significant finding was that the biosynthetic pathway encodes two regiospecific prenyltransferases, catalyzing the C- and N-terminal prenylation of muscoride. An antimicrobial activity screening showed that muscoride B had antimicrobial activity against *Bacillus cereus*.

Here I report the discovery of the muscoride biosynthetic pathway and the discovery of a novel antimicrobial peptide from cyanobacteria through genome mining. The results show that the variant is a novel muscoride, a linear bis-prenylated polyoxazole pentapeptide with antimicrobial activity.

Avainsanat – Nyckelord – Keywords

Antimicrobial peptides, cyanobacteria, natural products, ribosomally synthesized and post-translationally modified peptides

Säilytyspaikka – Förvaringställe – Where deposited

HELDA – Digital Repository of the University of Helsinki, E-thesis

Muita tietoja – Övriga uppgifter – Additional information

The thesis was supervised by Docent David Fewer.

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty	Laitos/Institution – Department	
Agrikultur-forstvetenskapliga fakulteten	Avdelningen för mikrobiologi	
Tekijä/Författare – Author		
Rose-Marie Andsten		
Työn nimi/Arbetets titel – Title		
Identifiering av biosyntesvägen för muscorid leder till upptäckt av en ny antimikrobiell naturprodukt		
Oppiaine/Läroämne – Subject		
Mikrobiologi		
Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/Sidoantal – Number of pages
Magisteravhandling	Juni 2020	60 + 8 bilagor

Tiivistelmä/Referat – Abstract

Bakterier är en bra källa till naturprodukter med komplexa kemiska strukturer och olika biologiska aktivitet. Många har terapeutiska egenskaper och hälften av läkemedel i kliniskt bruk idag, är direkt eller indirekt härledda ur naturprodukter. Läkemedelsindustrin slutade att investera i läkemedelsutveckling från naturresurser på grund av uppfattade begränsningar i kemiskt utrymme och svårigheter att återupptäcka kända föreningar och erhålla tillräckliga mängder av naturprodukter för kliniska prövningar. Det finns nu ett förnyat intresse för naturprodukter som läkemedelskandidat som drives av tekniska framsteg inom genomsekvensering och analytisk kemi. Cyanobakterier producerar en mängd naturprodukter med terapeutisk potential. Muscorid A är en ovanlig peptidalkaloid producerad av en markbunden sötvattens cyanobakterie med rapporterad antimikrobiell aktivitet.

Syfte med denna studien var att karakterisera det biosyntetiska ursprunget och bioaktiviteten av muscorid A. Jag identifierade det 12.7 kb muscorid (*mus*) biosyntetiska genklustret från utkast genomet av *Nostoc* sp. PCC 7960 med bioinformatiska analyser. Det biosyntetiska genklustret *mus* kodade enzymer för heterocyklisering, oxidering och prenylering av en prekursorpeptid. Jämförande genomik identifierade det biosyntestiska genklustret *mus* från ett opublicerat utkast genom av *Nostoc* sp. UHCC 0398 som kodar för en ny muscorid. Denna nya muscorid, muscorid B, upptäcktes från *Nostoc* sp. UHCC 0398 baserat på denna analys. Muscorid B extraherades med fastfasextraktion och högupplösande vätskekromatografi och den kemiska strukturen bestämdes kombinerat med kärnmagnetisk resonansspektroskopi och masspektrometri. Dessutom studerades muscorid prenyltransferasens funktion och evolutionära historia. Ett signifikant fynd var att biosyntesvägen kodar för två regiospecifika prenyltransferas som katalyserar prenyleringen av N- och C-terminalen för muscorid. Screening av antimikrobiell aktivitet visade att muscorid B hade antimikrobiell aktivitet mot *Bacillus cereus*.

Här rapporterar jag upptäckten av biosyntesvägen för muscorid och fyndet av en ny antimikrobiell peptid från cyanobakterier genom genomutvinning (eng. genome mining). Resultaten visar att varianten är en ny muscorid, en linjär bis-prenylerad polyoxazol pentapeptid med antimikrobiell aktivitet.

Avainsanat – Nyckelord – Keywords

Antimikrobiella peptider, cyanobakterier, naturprodukter, ribosomalt syntetiserade och posttranslationellt modifierade peptider

Säilytyspaikka – Förvaringställe – Where deposited

HELDA - Helsingfors universitets digitala publikationsarkiv, E-thesis

Muita tietoja – Övriga uppgifter – Additional information

Handledaren för avhandlingen var Docent David Fewer.

TABLE OF CONTENTS

ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
CD	Conserved domain
COSY	Correlation spectroscopy
DMAPP	Dimethylallyl pyrophosphate
EASY	Efficient adiabatic symmetrized
HAMBI	University of Helsinki microbial culture collection
HDMS	High definition mass spectrometry
HMBC	Heteronuclear multiple bond correlation
HPLC	High performance liquid chromatography
IAM	Institute of Applied Microbiology, Culture Collection, University of Tokyo
LC-MS	Liquid chromatography-mass spectrometry
MHA	Müller-Hinton agar
MS	Mass spectrometry
MUSCLE	Multiple sequence comparison by log-expectation
NCBI	National Center for Biotechnology Information
NMR	Nuclear magnetic resonance
PCC	The Pasteur Culture collection of Cyanobacteria
Q-TOF	Quadrupole time-of-flight mass spectrometer
RE	Recognition sequences
RiPP	Ribosomally synthesized and post-translationally modified peptide
ROESY	Rotating frame Overhauser effect spectroscopy
SPE	Solid phase extraction
TOCSY	Total correlation spectroscopy
TSA	Trypticase soy agar
TGY	Tryptone glucose yeast agar
UHCC	University of Helsinki Culture collection of Cyanobacteria
UPLC	Ultra performance liquid chromatography
UV	Ultraviolet
YM	Yeast and mold agar

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1 INTRODUCTION

Natural products are usually low-molecular-weight compounds synthesized by plants, animals and microorganisms through diverse biosynthetic pathways (Katz & Blatz 2016, Newman & Cragg 2020). They have complex chemical structures and different biological effects and represent various compound classes (Katz & Blatz 2016, Newman & Cragg 2020). The chemical diversity, stability and desirable bioactivities make these compounds interesting for pharmaceutical industry, biomedicine, agriculture, biotechnological applications and well as synthetic chemistry (Katz & Blatz 2016, Newman & Cragg 2020). Many of these compounds are synthesize by microbes and used in the pharmacological industry as lead compounds for drugs or used directly as semi synthetic derivatives (Katz & Blatz 2016, Newman & Cragg 2020). The technological developments in screening methods and genome mining has improved the identification and characterization of natural products and increased the interest among research (Katz & Blatz 2016).

Cyanobacteria produce a variety of compounds with therapeutic potential (Sivonen et al. 2010, Dittmann et al. 2015). Cyanobactins are a large group of cyanobacterial natural products that include linear and cyclic peptides with anticancer, multidrug reversing and antimicrobial activities (Sivonen et al. 2010, Leikoski et al. 2013). They are ribosomally synthesized and post-translationally modified and their biosynthetic genes are widespread among cyanobacteria (Sivonen et al. 2010, Arnison et al. 2013). The cyanobactin biosynthetic pathways are promiscuous with a broad substrate tolerance which makes them good platforms for genetic engineering (Sardar & Schmidt 2016, Gu & Schmidt 2017)

Many novel and known compounds with unidentified biosynthetic origin have been characterized as cyanobactins following the discovery of the cyanobactin biosynthetic pathway (Sivonen et al. 2010, Leikoski et al. 2013). Muscoride A is a linear peptide alkaloid produced by *Nostoc muscorum* IAM M-14 with a chemical structure that, suggest it may be synthesized through the cyanobactin biosynthetic pathway (Leikoski et al. 2013). It has an unusual linear structure with two contagious methyloxazoles and prenyl groups protecting the N- and C-terminal ends (Nagatsu et al. 1995). It is reported that, prenyl groups can increase interaction and access through the lipophilic membrane and oxazole containing natural products possess antimicrobial and cytotoxic activity (Botta et

al. 2005, Mhlongo et al. 2020). The rare chemical structure and potential antimicrobial activity makes polyoxazoles such as muscoride A an interesting target for research. Many oxazole containing natural products have been synthesized, including muscoride A (Wipf & Venkatraman 1996, Bagley et al. 1998, Muir et al. 1998, Coqueron et al. 2003, Yeh 2004, Amaike et al. 2012, Correa et al. 2013). The muscoride biosynthetic pathway has not been reported despite the interest in this compound from the organic chemistry community. The aim of this study was to characterize the biosynthetic origin and biological activity of muscoride A.

2 LITERATURE REVIEW

2.1 Cyanobacteria

Cyanobacteria are oxygenic photosynthetic bacteria that represents a morphologically, physiologically and ecologically diverse phylum (Castenholz 2015). This ancient group of prokaryotes is thought to have played a significant role in the history of life on Earth (Schopf 2012). Microfossils of ancestors to cyanobacteria date back to the Precambrian and it is believed that cyanobacteria impacted the oxygenation of the atmosphere around 2.45 billion years ago (Knoll 2008, Schopf 2012). Still today cyanobacteria are important organisms in the biogeochemical cycles and as primary producers (Whitton & Potts 2012).

Cyanobacteria can be unicellular, colonial or filamentous and some species can differentiate vegetative cells into atmospheric nitrogen fixating heterocysts or environmental stress tolerant resting cells, akinetes (Castenholz 2015). Cyanobacteria are found in diverse types of habitats as free-living or symbionts and they have many features that benefit and protect them against the surrounding environment (Whitton & Potts 2012). Some can tolerate extreme conditions such as high or freezing temperature, free sulfide, high levels of ultra-violet irradiation, high salinity and desiccation (Whitton & Potts 2012). They might have a polysaccharides or polypeptide sheath on the cell surface with protective UV-absorbing pigment, reserve nutrients and enzymes in their cytoplasm, gas vacuoles for buoyancy regulation or produce bioactive compounds (Walsby 1994, Castenholz 2015, Dittmann et al. 2015).

2.2 Bioactive secondary metabolites produced by cyanobacteria

Natural products are considered as secondary metabolites that are not directly involved in the growth and development of organisms but have biological activities that confer a competitive advantage against other organisms (Singh et al. 2017). Cyanobacteria synthesize a variety of bioactive secondary metabolites (Dittmann et al. 2015). Both aquatic and terrestrial cyanobacteria are known to produce secondary metabolites and in 2015 Dittmann et al. reported that over 1100 secondary metabolites have been found from 39 genera, where the majority belonged to *Lyngbya*, *Microcystis*, *Nostoc* and *Hapalosiphon*.

Species and strains, that occur in dense blooms are well studied due to their harmfulness to human, other mammals, birds, fish and invertebrates (Carmichael 1992, Burja et al. 2001, Paerl et al. 2001, Sivonen 2009). They produce compounds that can affect the odor and taste of drinking water and fish, or neuro-, hepato-, dermato- and cytotoxins that can cause human health problems, sickness or death of livestock, wildlife and pets (Carmichael 1992, Paerl et al. 2001, Suurnäkki et al. 2015, Sivonen 2009). Cyanobacteria also produce compounds that act against prokaryotic organisms, viruses, fungi and mammalian cell lines (Carmichael 1992). These metabolites target specific proteins and macromolecules and have antibacterial, anticancer, multidrug-reversing, anti-inflammatory, antifungal and antiviral activity (Burja et al. 2001, Welker & von Döhren 2006, Sivonen 2009, Dittmann et al. 2015).

A majority of the known cyanobacterial secondary metabolites are cyclic or linear peptides or lipopeptides, but also other chemical structural classes are represented such as lipids, alkaloids, polyketides and terpenes (Burja et al. 2001, Welker & von Döhren 2006). They are synthesized through different biosynthetic pathways producing structurally and functionally unique compounds (Dittmann et al. 2015). These features make them interesting and potential for biotechnological and pharmacological use and could serve as drug leads (Tan 2013).

2.3 Ribosomally synthesized and post-translationally modified peptides

One class of natural products are ribosomally synthesized and post-translationally modified peptides (RiPPs) produced in all domains of life (Arnison et al. 2013). They are peptides below 10 kDa and subdivided based on their chemical structures and biosynthetic mechanisms representing over 20 compound classes (Arnison et al. 2013). Though variation in their biosynthesis they follow the same basic logic in their pathways (Arnison et al. 2013). At first the gene cluster encoding a precursor peptide and the modification enzymes is translated on the ribosome (Arnison et al. 2013). The precursor peptide consists of one or several leader peptides, recognition sequences and core peptides (Arnison et al. 2013). The core peptide is the amino acid backbone for the mature peptide formed after enzymatic primary and tailoring modifications (Arnison et al. 2013, Funk & van der Donk 2017). Some subfamilies use same modification enzymes and common modifications are macrocyclization, N- and C- terminal modifications, epimerization, methylations, hydroxylation, disulfides, prenylation and proteolytic removal of the leader

peptide (Arnison et al. 2013). The modifications improve metabolic and structural stability and biological activity of the peptides (Arnison et al. 2013). Many RiPPs display antimicrobial activity and diverse growth inhibiting and cell membrane disruptive mechanisms have been reported (Arnison et al. 2013).

2.4 Cyanobactins

In 1980, two small cytotoxic cyclic peptides, ulicyclamide and ulithiacyclamide, were described from the ascidian *Lissoclinum patella* (Ireland & Scheuer 1980). At the time it was known that didemnid ascidians harbor symbiotic unicellular algae, but some understanding was lacking about the host, symbiont and their relationship (Ireland & Scheuer 1980). Later, in 1996, Sings and Rinehart hypothesized that the metabolic origin of these cyclic peptides might be the symbiotic cyanobacterium *Prochloron* sp. harboring *L. patella* (Sings & Rinehart 1996). Almost a decade later, it was proven by heterologous expression that *Prochloron didemni* synthetizes these peptides ribosomally and the patellamide biosynthetic pathway was described (Long et al. 2005, Schmidt et al. 2005). Several cyclic peptides from free-living and symbiotic cyanobacteria synthesized by a patellamide-like biosynthetic pathway were reported and in 2008, a global assembly line was introduced for these peptides and they were named cyanobactins (Donia et al. 2008).

Today, cyanobactins are defined as ribosomally synthesized and post-translationally modified cyclic or linear peptides produced by cyanobacteria (Arnison et al. 2013, Leikoski et al. 2013). They are a large compound class in cyanobacteria with diverse chemical structures and bioactivities such as antibacterial, antiviral, anticancer, antifungal (Donia & Schmidt 2010, Sivonen et al. 2010). The peptides are 3 to 20 amino acids long and have chemical structures such as oxazoles, oxazolines, thiazoles, thiazolines, disulfide bridges, prenyl, geranyl and methyl groups (Arnison et al. 2013, Leikoski et al. 2010, 2013). The cyanobactin biosynthetic genes are widespread and found in diverse cyanobacteria *Anabaena*, *Lyngbya*, *Microcystis*, *Nostoc*, *Prochloron* and *Trichodesminum* (Leikoski et al. 2009, Sivonen et al. 2010). It is estimated that the pathway occurs in 10 to 30% of cyanobacteria and in 2013 Leikoski *et al.* identified 31 gene clusters from 126 cyanobacteria genomes (Leikoski et al. 2009). The highly conserved biosynthetic genes and technological developments within genome sequencing, heterologous expression and databases has made the detection of

biosynthetic gene clusters easier and novel cyanobactins are characterized constantly (Velásquez & Van der Donk 2011, Leikoski et al. 2013, Dittmann et al. 2015).

2.5 The cyanobactin biosynthetic pathway

Cyanobactins are produced through the post-ribosomal peptide synthesis and the post-translational modifications are catalyzed by enzymes that modify the core sequence in the precursor peptide into the mature cyanobactin (Arnison et al. 2013). According to Leikoski et al. (2013) the cyanobactin gene cluster ranges from 8 to 19 kb and can encode between 8 to 24 genes. The nomenclature of the cyanobactin biosynthetic genes follow the first characterized cyanobactin patellamide pathway (*pat*) and each gene is designated with a letter (A-G) (Figure 1) (Schmidt et al. 2005, Arnison et al. 2013).

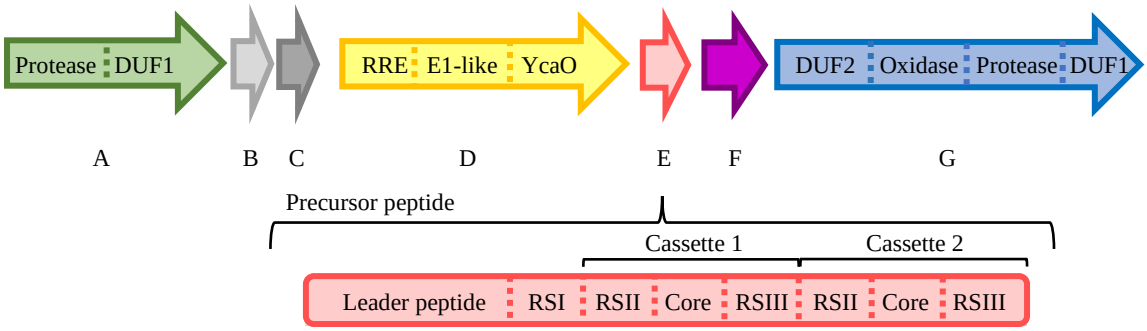


Figure 1. A general organization of the cyanobactin gene cluster with the biosynthetic genes (A-G) and the organization of the precursor peptide (E) leader, core and recognition sequence (RS) elements. The genes can be in a different order and the number of core cassettes can vary from one to four (figure modified from Gu et al. 2018a).

The translated precursor peptide (E) consists of a conserved leader peptide and the core sequence flanked by conserved recognition sequences I-III (RSI-III) (Figure 1). The core sequence encodes the amino acid backbone for the mature cyanobactin (Schmidt et al. 2005, Donia et al. 2006). One gene cluster can encode up to 10 precursor peptides producing the same or different cyanobactins (Donia & Schmidt 2011, Leikoski et al. 2013). Also, the hypervariable core sequences can vary from one to four, expanding the diversity of cyanobactins even more (Donia et al. 2006, Leikoski et al. 2009, 2013, Donia

& Schmidt 2011). The recognition sequences direct the modification enzymes (Donia et al. 2006, Sardar et al. 2015). If the precursor peptide encodes several core sequences flanked by recognition sequences, the area is called a cassette (Figure 1) (Gu et al. 2018b).

If the cyanobactin contains heterocycles, the first modification after the translation is the formation of heterocycles by a cyclodehydratase (D) on the core sequence (Figure 1) (McIntosh & Schmidt 2010, McIntosh et al. 2010a). It is followed by the proteolytic cleavage of the N- and C-termini by two different proteases (A and G) (Figure 1) (Lee et al. 2009). The macrocyclization of cyclic cyanobactins occurs in tandem with the cleavage of the C-termini and other modifications reactions such as oxidation, prenylation, geranylation and methylation may follow this (Lee et al. 2009, Leikoski et al. 2010, 2013, McIntosh et al. 2010b, Donia & Schmidt 2011).

Proteins and domains of unknown function are common in the cyanobactin biosynthetic pathway (Arnison et al. 2013). Proteins A and G have homologous C-terminal domains of unknown function (DUF1) and protein G an N-terminal domain of unknown function (DUF2) (Figure 1) (Agarwal et al. 2012, Koehnke et al. 2012). Two short proteins, B and C, are also commonly found in cyanobactin gene clusters but their functions are unknown (Donia et al. 2006). Though these genes were already described in 2005 from the patellamide pathway, their purpose still remains unknown (Schmidt et al. 2005).

2.5.1 Heterocyclization

The first modification of the core sequence is the formation of heterocycles, if the heterocyclase is present in the gene cluster (McIntosh & Schmidt 2010, McIntosh et al. 2010a). The D protein is a three-domain protein with a RiPP precursor peptide recognition element (RRE), an E1-like docking scaffold and an YcaO-like cyclodehydratase (Figure 1) (Dunbar et al. 2012, 2014, Burkhart et al. 2015). The RSI on the precursor peptide directs and interacts with the RRE and the ATP dependent reaction is catalyzed by the YcaO-like cyclodehydratase (McIntosh & Schmidt 2010, Dunbar et al. 2012, 2014, Burkhart et al. 2015, Sardar et al. 2015). Heterocycles oxazolines and thiazolines are derived from cysteine, serine or threonine residues in the reaction (McIntosh & Schmidt 2010, McIntosh et al. 2010a, Sardar et al. 2015).

2.5.2 Proteolytic cleavage and macrocyclization

After the heterocyclization the precursor peptide is proteolytically cleaved by two proteases (A and G) at the amino and carboxyl termini (Figure 1) (Schmidt et al. 2005, Lee et al. 2009). The A protein contains two domains, the subtilisin-like serine protease domain and a C-terminal domain of unknown function (DUF1) (Figure 1) (Lee et al. 2009). The RSII directs the protease A that catalyzes the cleavage of the N-terminus of the precursor peptide, removing the leader peptide (Lee et al. 2009, Sardar et al. 2015). The second protease (G) is also a subtilisin-like serine protease found in the multidomain protein G, harboring an oxidase domain and two DUFs (N- and C-terminal) (Figure 1) (Lee et al. 2009, Koehnke et al. 2012). The RSIII directs the protease G that cleaves the C-terminus at an azol(in)e heterocyclase or amino acid proline on the precursor peptide (Lee et al. 2009, Sardar et al. 2015). Protease G is also responsible for the N-C macrocyclization of cyclic cyanobactins, that occurs in tandem with the cleavage of the C-terminus (Lee et al. 2009, McIntosh et al. 2010b, Koehnke et al. 2012). When linear peptides are synthesized, protease G hydrolyses the C-terminus (Sardar et al. 2017).

2.5.3 Prenylation and geranylation

Both linear and cyclic prenylated cyanobactins are known and prenyltransferases (F) are found in many cyanobactin gene clusters, whereas some are active and others inactive (Figure 1) (Donia et al. 2006, Sivonen et al. 2010, Bent et al. 2013, Leikoski et al. 2010, 2013). Cyclic peptides are prenylated after the macrocyclization and linear peptides after the proteolytic cleavage (McIntosh et al. 2011, Leikoski et al. 2013, Sardar et al. 2017). There is no known recognition sequence for prenyltransferases (McIntosh et al. 2011, Hao et al. 2016). The prenylation occur on serine, threonine, tyrosine, tryptophan residue or the N-terminus of the peptide (McIntosh et al. 2011, Leikoski et al. 2012, 2013, Parajuli et al. 2016, Sardar et al. 2017). Prenyltransferase are ABBA-like proteins that catalyzes the reaction by adding a prenyl group (C_5) from the dimethylallyl pyrophosphate (DMAPP) substrate in forward or reverse orientation (McIntosh et al. 2011). C-prenylated cyanobactins can also form via Claisen rearrangement (McIntosh et al. 2011). Instead of the addition of a prenyl group, the F enzyme has been characterized to catalyze the forward O-geranylation of tyrosine by adding a geranyl group (C_{10}) (Leikoski et al 2012, Morita et al. 2018).

2.5.4 Other modifications

Heterocycles thiazoline and oxazoline can be oxidized into thiazole and oxazole, respectively (Sivonen et al. 2010). If the dehydrogenase is present in the gene cluster, it is usually encoded as a fused domain in the G protein, but sometimes found as a separate protein (Figure 1) (Schmidt et al. 2005, Donia & Schmidt 2011). The oxidase has two RREs at the N-terminus and a flavin mononucleotide-dependent nitroreductase at the C-terminal domain (Bent et al. 2016, Gao et al. 2018). The reaction is not leader dependent (Gao et al. 2018).

Only a few methylated cyanobactins are known though the methyltransferase domain is present in several gene clusters (Donia & Schmidt 2011, Leikoski et al. 2013, Sardar et al. 2017). In 2013 Leikoski et al. discovered several methyltransferase domains fused with A, F and G proteins by genome mining and reported the C-terminal methylated linear cyanobactins viridisamide A, aeruginosamide B and C (Figure 2). Aeruginosamide B is methylated by a SAM-dependent methyltransferase encoded in a didomain protein with a prenyltransferase (Leikoski et al. 2013, Sardar et al. 2017). The N-methylation has been described for a cyclic cyanobactin, microcyclamide and lately for a linear cyanobactin, scytodecamide (Ziemert et al. 2008, Crnkovic et al. 2020). The putative methyltransferase for scytodecamide is found in a didomain fused with a DUF (Crnkovic et al. 2020).

2.6 Linear cyanobactins

Linear cyanobactins were first described by Leikoski et al. in 2013 when three linear cyanobactins with N-prenylated and O-methylated termini were identified. Tetra- and pentapeptide aeruginosamide B (FFPC, core) and C (FFPVC, core) from *Microcystis aeruginosa* PCC 9432 and tripeptide viridisamide A (FIC, core) from *Oscillatoria nigro-viridis* PCC 7112 (Figure 2) (Leikoski et al. 2013). The gene clusters encoded N- and C-terminal proteases, heterocyclase and oxidase enzymes, putative methyltransferases and prenyltransferases and proteins of unknown function (Leikoski et al. 2013). The multicore precursor peptides had conserved leader sequences (LAELSEE, RSI) and cleavage sites (Leikoski et al. 2013). Based on the findings Leikoski et al. (2013) was able to define a previously reported linear peptide, aeruginosmaide A (IVPC, core), as a cyanobactin (Figure 2) (Lawton et al. 1999). It was also discussed that other earlier reported linear peptides from cyanobacteria, with unknown biosynthetic origins, might be synthesized by the cyanobactin biosynthetic pathway such as virenamide A (FVVC, core) from

Didemnid ascidian *Diplosoma virens* with decarboxylated C-terminus and prenylated N-terminus (Figure 2) (Carroll et al. 1996, Leikoski et al. 2013). Recently, scytodecamide, a novel linear cyanobactin, with N-terminal methylation and C-terminal amidation was described from *Scytonema* sp. UIC 10036 (Crnkovic et al. 2020). This decapeptide is produced by a cyanobactin-like gene cluster and showed cytotoxic activity (Crnkovic et al. 2020).

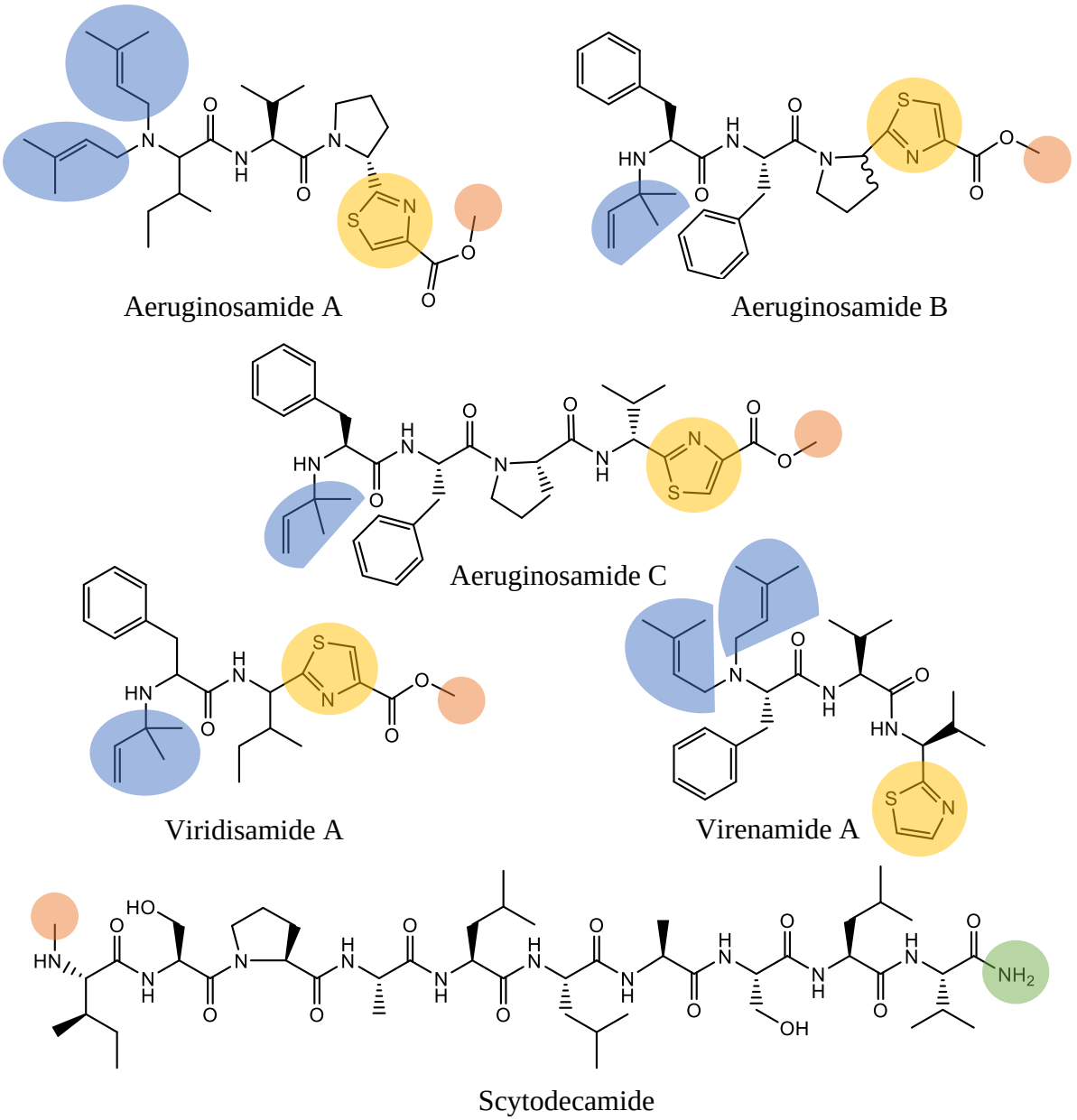


Figure 2. Linear natural products from cyanobacteria with prenylated (highlighted in blue), methylated (highlighted in red), amidated (highlighted in green) and thiazole (highlighted in yellow) structures. Chemical structures re-drawn using PerkinElmer

ChemDraw Professional 19.1 software based on the work of Carroll et al. 1996, Lawton et al. 1999, Leikoski et al. 2013 and Crnkovic et al. 2020.

2.7 Muscoride A

In 2013, Leikoski et al. also suggested that muscoride A could belong to the cyanobactin family. Muscoride A is a linear tetrapeptide alkaloid isolated from terrestrial freshwater cyanobacterium *Nostoc muscorum* IAM M-14 (Nagatsu et al. 1995). The peptide contains two contiguous methyloxazoles and a reverse prenyl group at the amino-terminus and a forward prenyl group at the carboxyl-terminus (Figure 3) (Nagatsu et al. 1995). Coupled heterocyclized amino acids are rare in nature and the total synthesis of muscoride A has been used for synthesis of oxazole and bisoxazole (Wipf & Venkatraman 1996, Muir et al. 1998, Bagley et al. 1998, Coqueron et al. 2003, Amaike et al. 2012, Correa et al. 2013). Polyoxazoles are known to possess antimicrobial activity and an agar well method-based assay showed that muscoride A had antimicrobial activity against *Bacillus subtilis* (3-6 mm) and no activity against *Escherichia coli* (Nagatsu et al. 1995, Molohon et al. 2011).

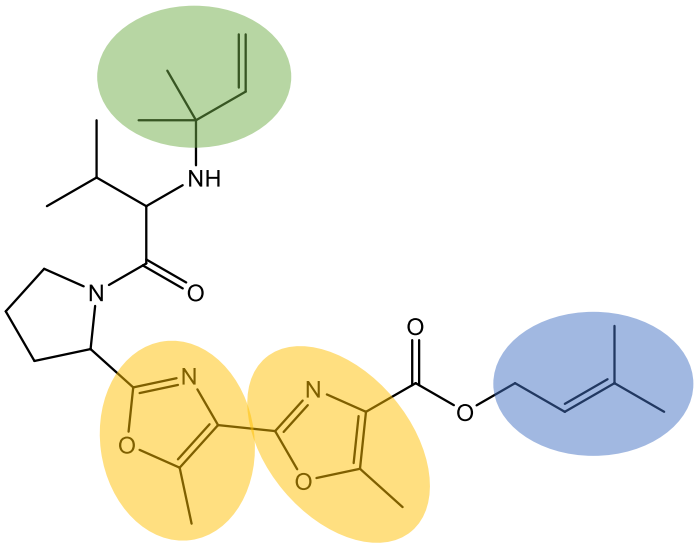


Figure 3. The chemical structure of muscoride A produced by *Nostoc muscorum* IAM M-14. A reverse prenyl group at the amino-terminus (highlighted in green), a forward prenyl group at the carboxyl-terminus (highlighted in blue) and two contiguous methyloxazoles (highlighted in yellow). Chemical structure re-drawn using PerkinElmer ChemDraw Professional 19.1 software based on the work of Nagatsu et al. 1995.

3 RESEARCH AIMS

The chemical structure of muscoride A is rare and it has been synthesized several times by organic chemists to yield oxazole and bisoxazoles (Wipf & Venkatraman 1996, Muir et al. 1998, Bagley et al. 1998, Coqueron et al. 2003, Amaike et al. 2012, Correa et al. 2013). Despite this the biosynthetic origin remains unknown. It has been suggested that muscoride A is synthesized through the post-translational modification of a precursor protein via the cyanobactin biosynthetic pathway (Leikoski et al. 2013). The aim of this study was to test this hypothesis and determining the biological origins and bioactivity of muscoride. The specific objectives of this thesis were to:

i. Identify the muscoride biosynthetic pathway
ii. Catalog genetic and chemical diversity of muscoride pathways
iii. Develop a biosynthetic scheme for the production of muscorides
iv. Determine the antimicrobial activity of muscorides

4 MATERIALS AND METHODS

4.1 Identification and annotation of the cyanobactin biosynthetic gene cluster from *Nostoc* sp. PCC 7906

An identical cyanobacterial pure strain to *Nostoc muscorum* IAM M-14, *Nostoc* sp. PCC 7906 obtained from the Pasteur Culture collection of Cyanobacteria (PCC) had been shown to produce muscoride A and high-molecular-weight genomic DNA of the strain had been extracted, sequenced and previously assembled before (Mattila et al. 2019). The biosynthetic gene cluster from *Nostoc* sp. PCC 7906 was identified using the N-terminal protease AgeA (accession number CCH92969) and the C-terminal protease AgeG (accession number CCH92964) from the aeruginosamide pathway (Leikoski et al. 2013) as query sequences against tBLASTn in the National Center of Biotechnology Information (NCBI) database. The gene cluster was visualized using Artemis (ACT, Artemis comparison tool) (Carver et al. 2005) and the genes were manually annotated using the NCBI Conserved Domain (CD) Database searches. After the cyanobactin gene cluster from *Nostoc* sp. PCC 7906 was detected, tBLASTn searches were made against the non-redundant database at NCBI and 70 unpublished draft genomes from (data not shown). The annotation of the cyanobactin gene cluster from *Nostoc* sp. UHCC 0398 was performed as previously described for *Nostoc* sp. PCC 7096. High-molecular-weight genomic DNA of the strain had been extracted, sequenced and previously assembled before (Mattila et al. 2019).

4.2 Phylogenetic analysis of prenyltransferase enzymes

Phylogenetic analysis and the prenyltransferase phylogenetic tree was constructed with LIRMM (Dereeper et al. 2008) using the MusF1 (accession number QGA72590) and MusF2 (accession number QGA72591) prenyltransferase from *Nostoc* sp. PCC 7906. These enzymes were as a query in BLASTp to find other prenyltransferases sequences. The prenyltransferase sequences producing significant alignments (68) were downloaded and aligned. This set of prenyltransferase sequences were submitted to LIRMM analysis using the default mode. The alignment analysis was made by multiple sequence comparison by log-expectation (MUSCLE) version 3.5 (Edgar 2004), curation by Gblocks version 0.91b (Castresana J. 2000), phylogeny by PhyML 3.0: new algorithms,

methods and utilities (Guindon et al. 2010) and the tree drawing with TreeDyn version May 2006. The resulting phylogenetic tree was rooted at the mid-point.

4.3 Liquid chromatography/mass spectrometry guided isolation of the muscoride variant from *Nostoc* sp. UHCC 0398

4.3.1 Cultivation of *Nostoc* sp. UHCC 0398

Strain *Nostoc* sp. UHCC 0398 obtained from the University of Helsinki Culture collection of Cyanobacteria (UHCC) was cultivated for chemical analysis and antimicrobial activity screening in 36 L (12 batches, 3 L culture bottles) of Z8 medium (Kótai 1972) in constant light with photon irradiance of 5 to 10 µmol m^{-2} s^{-1} with aeration at 20°C to 25°C for 15 days. The cells were harvested by gravitation (24 h) and centrifugation (Sorvall LYNX 6000 Superspeed Centrifuge, Thermo Fisher Scientific, MA, USA) at 7 000 × *g* for 7 min at 20°C. The collected cells were freeze-dried (LyoCube 4-8 LSC*plus* & beta 2-8 LSC*plus*, Martin Christ, Germany) and the yield was 7.70 g of dried sample.

4.3.2 Chemical analysis

Freeze-dried *Nostoc* sp. UHCC 0398 cells (20 mg) were lysed in one ml of methanol (LiChrosolv®, Merck KGaA, Germany) in a two ml plastic tube with 200 µl of glass beads (Cell disruption media, 0.5 mm, Scientific Industries, Inc., NY, USA). The cells were homogenized with FastPrep®-24 (MP Biomedicals) for 45 s at 6.5 m s^{-1} and centrifuged at 20,000 × *g* for 5 min. A sample of 200 µl supernatant was filtered (pore size 0.2 µm, 13 mm Syringe Filter, PTFE membrane, VWR international, USA) into a glass vial. The extract was analyzed using ultra-performance liquid chromatography and high definition mass spectroscopy (UPLC-HDMS).

The analysis was performed with an ultra-performance liquid chromatography-high definition mass spectrometry system (ACQUITY UPLC-SYNAPT G2-Si HDMS, Waters, MA, USA) combined with a Kinetex® C8 LC column (50 × 2.1 mm, 1.7 µm, 100 Å, Phenomenex, CA, USA). The column was heated to 40°C and the sample flow rate was 0.3 ml min^{-1} with an injection volume of 1.0 µl. Solvents used in the analysis were 0.1% aqueous formic acid (HiPerSolv CHROMANORM® for LC/MS, VWR international, Belgium) (solvent A) and 1% formic acid in acetonitrile (HiPerSolv CHROMANORM®, HPLC-gradient grade, VWR international, France) / isopropanol

(1:1) (solvent B). A gradient program from 5% to 100% of solvent B during 7 min was used and the column was washed with 95% of solvent A for 2.5 min. Positive electrospray ionization mode was used at a scan range *m/z* 50 – 2000. The capillary voltage was 1.5 kV, source temperature 120°C, sampling cone 40.0, source offset 80.0, desolvation temperature 600°C, desolvation gas flow 1000 l h^{-1} and nebuliser gas flow 6.5 bar.

Identification of the muscoride variant from *Nostoc* sp. UHCC 0398 methanol crude extracts was based on the calculated theoretical molecular weight predicted from the core peptide amino acids sequence using PerkinElmer ChemDraw Professional software by Mr. Matti Wahlsten.

4.4 Extraction of the muscoride variant from *Nostoc* sp. UHCC 0398

4.4.1 Lysis of the cells

Two grams of lyophilized *Nostoc* sp. UHCC 0398 cells was weighed in two 50 ml plastic centrifuge tubes (1 g × 2). The cells were lysed in 30 ml of methanol (HiPerSolv CHROMANORM, VWR international, France) using a SilentCrusher M homogenizer (Heidolph Instruments GmbH & Co. KG, Germany) for 30 s at 15,000 rpm. The suspensions were centrifuged (Centrifuge 5804 R, Eppendorf AG, Germany) at 10,000 × *g* for 5 min at 22°C and the supernatants were collected in a round-bottom flask. The homogenization was repeated, and the total yield of supernatant was ~120 ml. To confirm that muscoride was present in the supernatant a sample was made for UPLC-HDMS analysis. The sample was prepared in a glass vial and 100 µl of supernatant was diluted in 100 µl of methanol. Octadecylsilyl silica gel (ODS) (5 g, 100 – 200 mesh, Chromtorex, Fuji-Davison Chemical Ltd., Japan) was added to the remaining supernatant and the sample was rotary evaporated (Rotavapor R-200, Vacuum Controller V-800, Heating Bath B-490, Büchi, Switzerland) to dryness at 250 mbar at 40°C and the dry sample was stored in 4°C.

4.4.2 Solid phase extraction

The silica gel bound sample was fractionated in two batches using solid phase extraction (SPE). A Strata® SI-1 Silica cartridge (5 g/20 ml, 55µm, 70Å, Giga Tubes Teflon, Phenomenex, CA, USA) was preconditioned with 20 ml of isopropanol (HiPerSolv CHROMANORM, VWR international, Belgium) as the conditioning solvent and 15 ml

of heptane (CHROMASOLV, Honeywell Riedel-de Haën®, Germany) as the equilibrant solvent. Half of the silica gel bound sample (2.5 g) was loaded onto the cartridge and washed with 25 ml of heptane, 25 ml of heptane-ethyl acetate (1:1), 25 ml of ethyl acetate (LiChrosolv, Merck KGaA, Germany) and 30 ml of methanol. The organic solvents were pipetted with glass pipettes and the eluates were collected by color in separate glass test tubes. Filtered compressed air was used during the extraction to pass the eluents through the sorbent. The rest of the dried sample in silica gel (2.5 g) was treated the same way using a new Strata® SI-1 Silica cartridge. To find out in which fractions contained muscoride, a 100 µl sample from each fraction was prepared in a glass vial for UPLC-HDMS analysis. The rest of the fractions were stored in 4°C.

4.4.3 Preparative high-performance liquid chromatography-mass spectroscopy

The fractions that contained muscoride were dried under a nitrogen gas stream and further purified with a high-performance liquid chromatography (HPLC) combined with a mass spectrometer (MS). The analysis was performed with Waters AutoPurification™ System (MA, USA) combined with Waters MassLynx™ software and Waters FractionLynx™ application manager. The mass spectra were recorded with Waters Micromass ZQ and the HPLC system consisted of a Waters 2767 Sample Manager, Waters UV Fraction manager, Waters System Fluidics Organizer, Waters Active Flow Splitter, Waters 2545 Binary Gradient Module, Waters 515 HPLC pump and a Waters 2996 Photodiode Array Detector. The fractions were collected by mass and the trigger was set to *m/z* 512 due to best intensity.

The sample was dissolved in 2 ml of eluent (30% acetonitrile/isopropanol (1:1), 70% 0.1% aqueous formic acid), sonicated (Branson 5510, MO, USA), centrifuged at 21,100 × *g* for 5 min and filtered (pore size 0.2 µm, 13 mm syringe filter, PTFE membrane, VWR international, USA). A Luna C8 (2) column (150 × 10.00 mm, 5 µm, Phenomenex, CA, USA) and electrospray ionization in positive ion mode was used during the analysis. The flow rate was 5 ml min^{-1} and a semi-preparative mode was used with an injection volume between 25 to 150 µl. The mobile phase consisted of 0.1% aqueous formic acid (solvent A) and acetonitrile/isopropanol (1:1) (solvent B). A gradient program from 30% to 100% of solvent B during 7 min was used and the total run time was 14 min. Various amounts of fractions were collected 6-7 min after run start and 63 glass tubes were collected in total.

The organic solvents were evaporated from the fractions under a nitrogen gas stream. Fractions left with water were partitioned with ethyl acetate/Milli-Q water (1:1). Ethyl acetate was added to each tube (1:1) and the tubes were vortexed. After letting the tubes rest for 30 to 60 s two layers were formed and the upper layers with ethyl acetate and muscoride were collected in new glass tubes. The procedure was repeated once and 23 glass tubes with 8-10 ml of ethyl acetate were collected in total. The samples were evaporated to dryness under a nitrogen gas stream, pooled with 3.5 ml of methanol and collected in a new glass tube. To confirm muscoride was in the sample, 1 µl of sample was diluted in 199 µl of methanol in a glass vial and analyzed with UPLC-HDMS.

According to the results from UPLC-HDMS analysis, muscoride was not pure enough and hence fractionated again using the previously described HPLC-MS system. The sample was evaporated to dryness under a nitrogen gas stream and the following alterations were made for the analysis. The dry sample was dissolved in 1 ml of 50% aqueous methanol (LiChrosolv®, Merck KGaA, Germany) solution and the injection volume was 150 µl. The mobile phase consisted of 0.1% aqueous formic acid solution (solvent A) and 50% aqueous methanol solution (solvent B). A gradient program from 50% to 80% of solvent B during 7 min was used and the total run time was 14 min. Eight fractions (~2.5 ml/tube) were collected and the organic solvents was evaporated under a nitrogen gas stream. The fractions were left with ~15 ml of water and they were concentrated using a Strata™-X 33 µm polymeric reversed phase cartridge (200 mg/6 ml, Phenomenex, CA, USA). The cartridge was preconditioned using 5 ml of methanol as conditioning solvent and 5 ml of Milli-Q water as equilibrate solvent. Loaded the sample onto the cartridge, washed it with 5 ml of Milli-Q water that was blown with compressed air. Used 5 ml methanol as elution solution. Measured 100 µl of the sample in a glass vial for UPLC-HDMS analysis and evaporated the rest of the sample to dryness under a nitrogen gas stream. The recovered muscoride was pure enough for structural analysis by nuclear magnetic resonance (NMR) spectroscopy.

4.5 Nuclear magnetic resonance spectroscopy

A structural analysis of muscoride purified from *Nostoc* sp. UHCC 0398 was done by NMR spectroscopy. The analysis was done at VERIFIN (Department of Chemistry, University of Helsinki, Kumpula, Finland) by Harri Koskela. A 20 µg muscoride sample was dissolved in 30 µl of pyridine-d_5 (Euriso-Top, Saint-Aubin Cedex, France) and

transferred into a 1.7 mm NMR tube (Bruker BioSpin, Rheinstetten, Germany). The analysis was done at 11.75 T using a Bruker Avance III 500 NMR spectrometer (Rheinstetten, Germany) combined with a 1.7 mm inverse z-gradient triple-resonance (^{1}H, ^{31}P, ^{13}C) microcoil probehead at 300 K. Assignments were made via proton nuclear magnetic resonance (^{1}H NMR), heteronuclear multiple bond correlation (HMBC, J_{LR} = 8 Hz), homonuclear correlation spectroscopy (COSY-45), homonuclear total correlation spectroscopy (TOCSY) and efficient adiabatic symmetrized rotating frame nuclear Overhauser enhancement spectroscopy (EASY ROESY). The characteristics for the assignments are shown in Table S1 in Appendix 1.

4.6 Prenylation assay

The *mus* biosynthetic gene cluster encodes two prenyltransferases, MusF1 and MusF2 (Figure 4). It was not clear from the sequence analysis which enzyme catalyzed the N- and O-prenylation of the muscoride A intermediate (data not shown). Therefore, an assay was performed using both synthetic amino and carboxyl substrates with the MusF1 and MusF2 prenyltransferase (Table 1). The prenyltransferase enzymes MusF1 (0.27 mM) and MusF2 (0.021 mM) from *Nostoc* sp. UHCC 0398 were already provided by Dr. Hideo Iwai (University of Helsinki). The substrates used in this assay were 5-methyl-2-phenyl-1,3-oxazole-4-carboxylic acid (10 mM, Sigma-Aldrich) and VPPP (1 mg/ml, 1M 408.49 g/l, GenScript).

Two different reaction mixtures incubated in two different temperatures (28°C and 37°C) were tested for both substrates (Table 1). All reactions were carried out in the same amount of $MgCl_2$ (50 mM), NaCl (1 M) and buffer (HEPES, 100 mM, pH 7.5) but differed in the concentration of enzyme (in a buffer containing 150 mM NaCl, 10 mM HEPES (pH 7.5) and 0.1 mM TCEP), dimethylallyl pyrophosphate (DMAPP, 5 mM), substrate and sterile Milli-Q water (Table 1). Negative controls were made for both amino and carboxyl substrates according to reaction mixture 1. without enzyme and incubated in 28°C (Table 1). The reagents were mixed in 1.5 ml plastic tubes and incubated for 48 h (Table 1). After the incubation the tubes were vortexed and 200 µl samples were filtered (pore size 0.2 µm, 13 mm Syringe Filter, PTFE membrane, VWR international, USA) into glass vials. Products of each reaction was analyzed by UPLC-HDMS.

Table 1. Reaction mixtures for the prenylation assays.

	Reaction mixture 1.		Reaction mixture 2.	
Reagent	Volume (µl)	Final concentration	Volume (µl)	Final concentration
$MgCl_2$ (50 mM)	48	12 mM	48	12 mM
NaCl (1 M)	30	150 mM	30	150 mM
HEPES buffer (100 mM)	20	10 mM	20	10 mM
DMAPP (5 mM)	40	1 mM	12	300 µM
Substrate amino (1 mg/ml, 1M 408.49 g/l)	8.2	100 µM	16.3	200 µM
Enzyme MusF2 (0.021 mM)	47.6	5 µM	95.2	50 µM
Milli-Q water	6.2			
Substrate carboxyl (10 mM)	2	100 µM	4	200 µM
Enzyme MusF1 (0.27 mM)	3.7	5 µM	37	50 µM
Milli-Q water	56.3		49	

4.7 Antimicrobial susceptibility test

The polyoxazole structure of muscoride A suggests antimicrobial activity. Hence strains *Nostoc* sp. UHCC 0398 and *Nostoc* sp. PCC 7906 were evaluated for antimicrobial activity. The antimicrobial assay was performed using two different diffusion methods, the disk diffusion method and the agar well diffusion method. Methanol crude extracts from *Nostoc* sp. UHCC 0398 and *Nostoc* sp. PCC 7906 and a semi-pure extract of muscoride from *Nostoc* sp. UHCC 0398 were tested against one yeast and different bacterial strains obtained from the University of Helsinki microbial culture collection (HAMBI) (Table 2). The organisms included in the assays were recovered from -80°C and grown on suitable media and temperature for 24 h before the assays were done (Table 2).

The disk diffusion assay was performed based on Kirby-Bauer disk diffusion susceptibility test protocol with some modifications (Bauer et al. 1966). The inoculums were prepared from the recovered organisms to be tested by picking four to five colonies with sterile loops (Table 2). The colonies were suspended in 2 ml of sterile Milli-Q water and mixed using a vortexer. Organism or sterile Milli-Q water was added until the suspensions reached a turbidity of 0.5 McFarland standard (Pro-Lab Diagnostics Inc., SD2300, 32211). The inoculums were evenly streaked on Müller-Hinton agar (MHA) plates with sterile cotton swabs.

Prepared the paper disks with the semi-pure muscoride extract from *Nostoc* sp. UHCC 0398. The sample was a collection of semi-pure fractions once fractionated using HPLC-MS. The dry sample (2.65 mg) was dissolved in 4 ml of methanol and 100 µl (66.25 µg/disk) of sample was infiltered into sterile blank paper disks (21-blank, Abtek Biologicals). Used methanol (50 µl/disk) as negative control, and kanamycin (1000 µg/disk, lot. IK12a, Abtek Biologicals Ltd. 11.2017) as positive control for bacteria and nystatin (150 µg/disk, 5 mg ml^{-1}, Sigma-Aldrich) as positive control for the yeast. The disks were placed on the inoculated agar surface with sterile forceps and the plates were incubated for 24 h in suitable temperatures (Table 2).

Table 2. Organisms included in the different antimicrobial assays with corresponding recovery plates, incubation plates and temperatures.

Organism (HAMBI nr.)	Recovery plates	Incubation temperature (°C)	Assay	Inoculum plate
Bacillus cereus (1881)	TSA	28	D, W	MHA
Bacillus subtilis (251)	TGY	28	D, W	MHA
Escherichia coli (396)	TGY	37	D	MHA
Micrococcus luteus (2688)	TGY	28	D	MHA
Pseudomonas aeruginosa (25)	TGY	37	D, W	MHA
Staphylococcus aureus (11)	TSA	37	D, W	MHA
Candida albicans (261)	YM	30	D	YM, MHA

D; disk diffusion method, W; agar well diffusion method

Bacteria used in the agar well diffusion method assay are listed in Table 2. The inoculums were made and applied on Müller-Hinton agar plates as described in the disk diffusion method. After the plates were inoculated, holes with a diameter of ~6 mm were made with sterile Pasteur pipets in the agar. The semi-pure *Nostoc* sp. UHCC 0398 muscoride sample (~3.3 ml) left from the disk diffusion assay was concentrated and the methanol was evaporated under a stream of gaseous nitrogen. Also, methanol crude extracts from *Nostoc* sp. PCC 7906 and four different cultivation batches of *Nostoc* sp. UHCC 0398 were tested. The methanol crude extracts were prepared from ~21 mg of freeze-dried cells as described in section 4.3.2 and dried under a nitrogen stream before storage in 4°C.

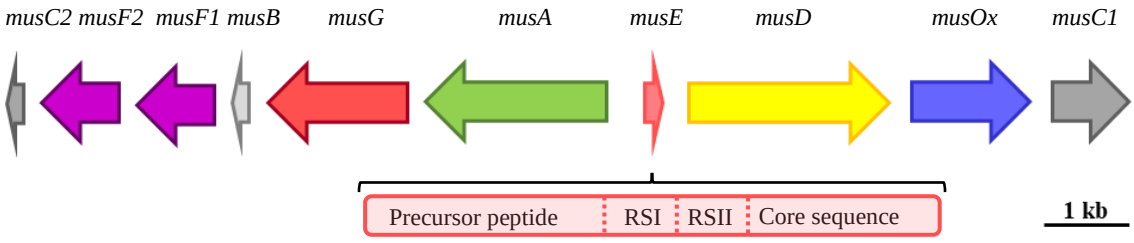
The dry semi-pure muscoride sample and crude extracts were dissolved in 200 µl of methanol, mixed with a vortexer and sonicated. The samples and controls were pipetted into the wells, 45 µl of sample, 40 µl of methanol as negative control and 5 µl of kanamycin (25 mg ml^{-1}) as positive control. The plates were incubated for 24 h in suitable temperatures listed in Table 2.

5 RESULTS

5.1 Identification of the muscoride gene cluster from *Nostoc* sp. PCC 7906

A 12.7 kb cyanobactin-like gene cluster encoding 10 proteins organized in a bidirectional operon was identified from *Nostoc* sp. PCC 7906 (Figure 4). The muscoride (*mus*) gene cluster comprised of an N-terminal protease (MusA), a heterocyclase (MusD), a precursor peptide (MusE), an oxidase (MusOx), a methyltransferase (MusG), two prenyltransferases (MusF1 and MusF2) and three unknown proteins (MusB, MusC1 and MusC2) (Figure 4, Table 3). The MusG protein lacked the C-terminal protease domain and the macrocyclase domain (Figure 4). The precursor peptide composed of 66 amino acids and one cassette with two conserved recognition sequences QLDLSEEEL (RSI) and GVSPS (RSII) and a single core sequence with four amino acids; Val, Pro, Thr, Thr (VPTT) (Figure 4, Table 3). The third recognition sequence (RSIII) was absent from the precursor peptide (Figure 4). Based on typical cyanobactin biosynthetic genes identified from the *mus* gene cluster, a biosynthetic scheme for the biosynthesis of muscoride A is suggested and presented in Figure 5.

A



B

```
PCC 7906          MKPENQKQTRLQPKLSEPTVRVMPGIQLDLSEEELTNGGVGVGGVNASQYGGGDGGVSPSVPTTGV
UHCC 0398         MKPENQKQTRLQPKLSEPTVRVMPGIQLDLSEEELTNGGVGVGGVNASQYGGGDGGVSPSIPSTSGV
SP3               MKPENQKQTRLQPKLSEPTVRVMPGIQLDLSEEELTNGGVGVGGVNASQYGGGDGGVSPSIPTSSGV
PCC 9307      MRPENHEENQKQTRLQPKLSEPTVRVMPGIQLDLSEEELTNGGVGVGGVNASQYGGGDGGVSPSIPTSSGV
HYL 09-2-6        MKPENQKQTRLQPKLSEPTVRVMPGIQLDLSEEELTNGGVGVGGVNASQYGGGDGGVSPSVPTSTI
Z1                MKPENQKQTRLQPKLSEPTVRVMPGIQLDLSEEELTNGGVGVGGVNASQYGGGDGGVSPSIPTSV
```

Figure 4. The *mus* biosynthetic pathway genes in panel A and precursor peptide sequences from different strains of *Nostoc*, showing the recognition sequences (underlined) and core sequences (highlighted in gray) in panel B.

The *mus* biosynthetic gene cluster was identified from 12 *Nostoc* strains, each of which encoded 10 biosynthetic genes with the same organization and with sequence identities ≥ 88% at nucleotide level compared to corresponding genes from the *Nostoc* sp. PCC 7906 *mus* biosynthetic gene cluster (Figure 4). The precursor peptides encoded the same recognition sequences RSI and RSII, but the core sequences had some variation encoding four to five amino acids (Figure 4). The core sequence of *Nostoc* sp. UHCC 0398 composed of five amino acids; Ile, Pro, Ser, Thr, Ser (IPSTS), indicating a muscoride variant with three contiguous oxazoles could be synthesized (Figure 4). The *mus* biosynthetic gene cluster with the 10 biosynthetic genes were encoded in *Nostoc* sp. UHCC 0398 (Table 3).

Table 3. Annotation of the *mus* gene clusters of *Nostoc* sp. PCC 7906 and *Nostoc* sp. UHCC 0398.

Nostoc sp. PCC 7906			NCBI CD-search results			
Protein	Length (aa)	Predicted function	Description	Identity (aa %)	Organism	Accession
MusC1	301	Unknown	Cyanobactin biosynthesis PatC/TenC/TruC family protein	84	*Nostoc linckia*	WP_099076588
MusOx	461	Oxidase	SagB/ThcOx family dehydrogenase	95	*Nostoc linckia*	WP_099068416
MusD	781	Heterocyclase	Adenylate cyclase	96	*Nostoc linckia*	WP_099068417
MusE	66	Precursor peptide	Hypothetical protein	96	*Nostoc linckia*	WP_099068418
MusA	706	N-terminal protease	PatA/PatG family cyanobactin maturation protease	93	*Nostoc linckia*	WP_099068419
MusG	544	Methyltransferase	PatA/PatG family cyanobactin maturation protease	95	*Nostoc linckia*	WP_099068420
MusB	66	Unknown	Cyanobactin biosynthesis PatB/AcyB/McaB family protein	100	*Nostoc linckia*	WP_099068421
MusF1	300	Prenyltransferase	LynF/TruF/PatF family peptide *O*-prenyltransferase	95	*Nostoc linckia*	WP_099072935
MusF2	303	Prenyltransferase	LynF/TruF/PatF family peptide *O*-prenyltransferase	66	*Nostoc linckia*	WP_099072934
MusC2	70	Unknown	Cyanobactin biosynthesis PatC/TenC/TruC family protein	90	*Nostoc linckia*	WP_099072933

Nostoc sp. UHCC 0398			NCBI CD-search results		
MusC1	308	Unknown	Cyanobactin biosynthesis PatC/TenC/TruC family protein	95	*Nostoc linckia* WP_099068415
MusOx	458	Oxidase	SagB/ThcOx family dehydrogenase	95	*Nostoc linckia* WP_099068416
MusD	781	Heterocyclase	Adenylate cyclase	93	*Nostoc linckia* WP_099068417
MusE	67	Precursor peptide	Hypothetical protein	96	*Nostoc linckia* WP_099068418
MusA	705	N-terminal protease	PatA/PatG family cyanobactin maturation protease	99	*Nostoc linckia* WP_099068419
MusG	544	Methyltransferase	PatA/PatG family cyanobactin maturation protease	97	*Nostoc linckia* WP_099068420
MusB	66	Unknown	Cyanobactin biosynthesis PatB/AcyB/McaB family protein	98	*Nostoc linckia* WP_099068421
MusF1	300	Prenyltransferase	LynF/TruF/PatF family peptide *O*-prenyltransferase	97	*Nostoc linckia* WP_099072935
MusF2	300	Prenyltransferase	LynF/TruF/PatF family peptide *O*-prenyltransferase	97	*Nostoc linckia* WP_099072934
MusC2	66	Unknown	Cyanobactin biosynthesis PatC/TenC/TruC family protein	98	*Nostoc linckia* WP_099072933

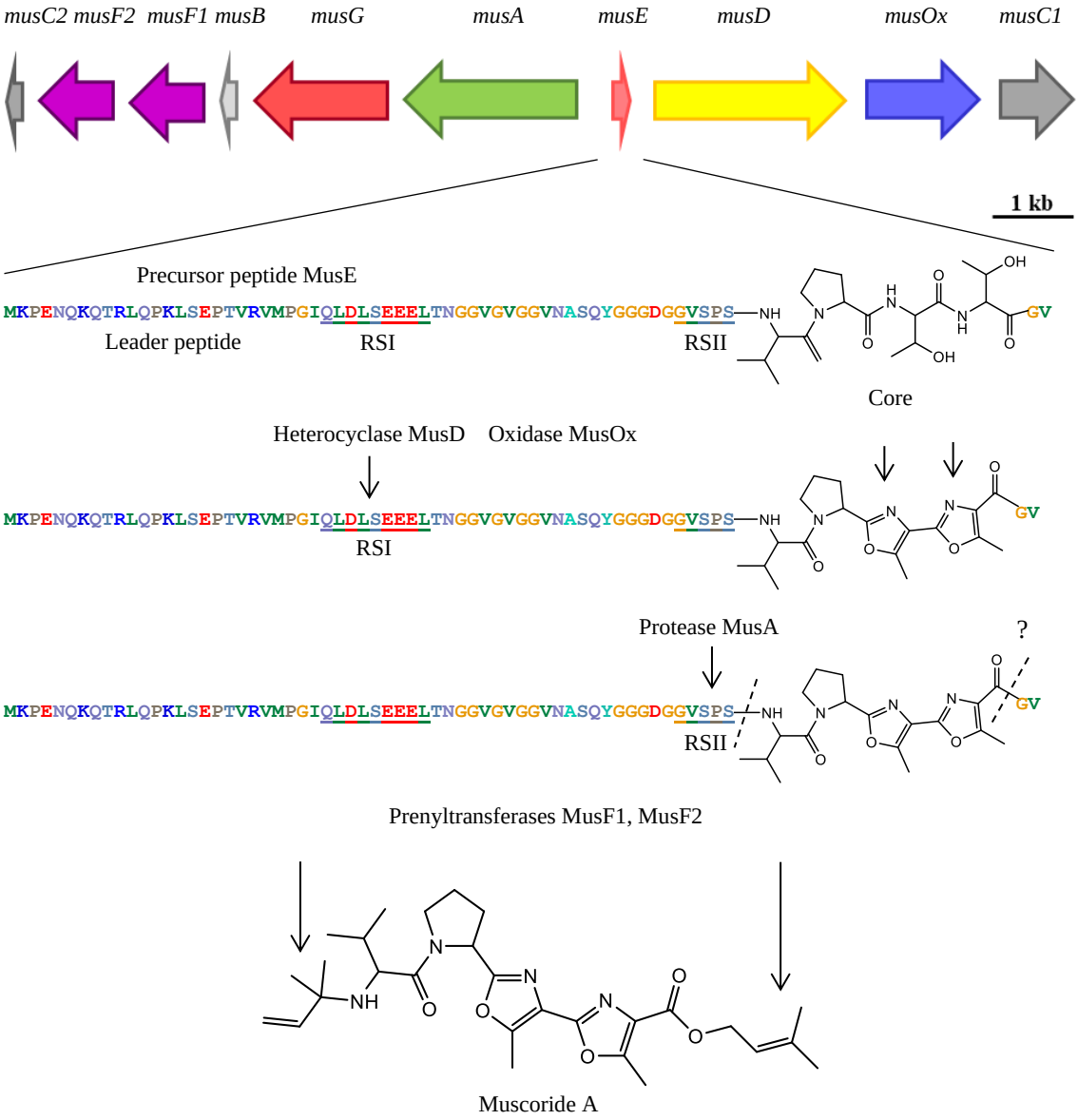


Figure 5. The muscoride (*mus*) biosynthetic gene cluster from *Nostoc* sp. PCC 7906 and a putative biosynthetic scheme for the biosynthesis of muscoride A.

5.2 Phylogenetic analysis of muscoride prenyltransferases

A phylogenetic tree was constructed from known cyanobactin prenyltransferases and the MusF1 and MusF2 prenyltransferases from *Nostoc* sp. UHCC 0398 and *Nostoc* sp. PCC 7906 (Figure 6). The MusF1 and MusF2 prenyltransferases fell into two distinct sister clades and they were distantly related to prenyltransferases VirF1 and AgeF1 (Figure 6).

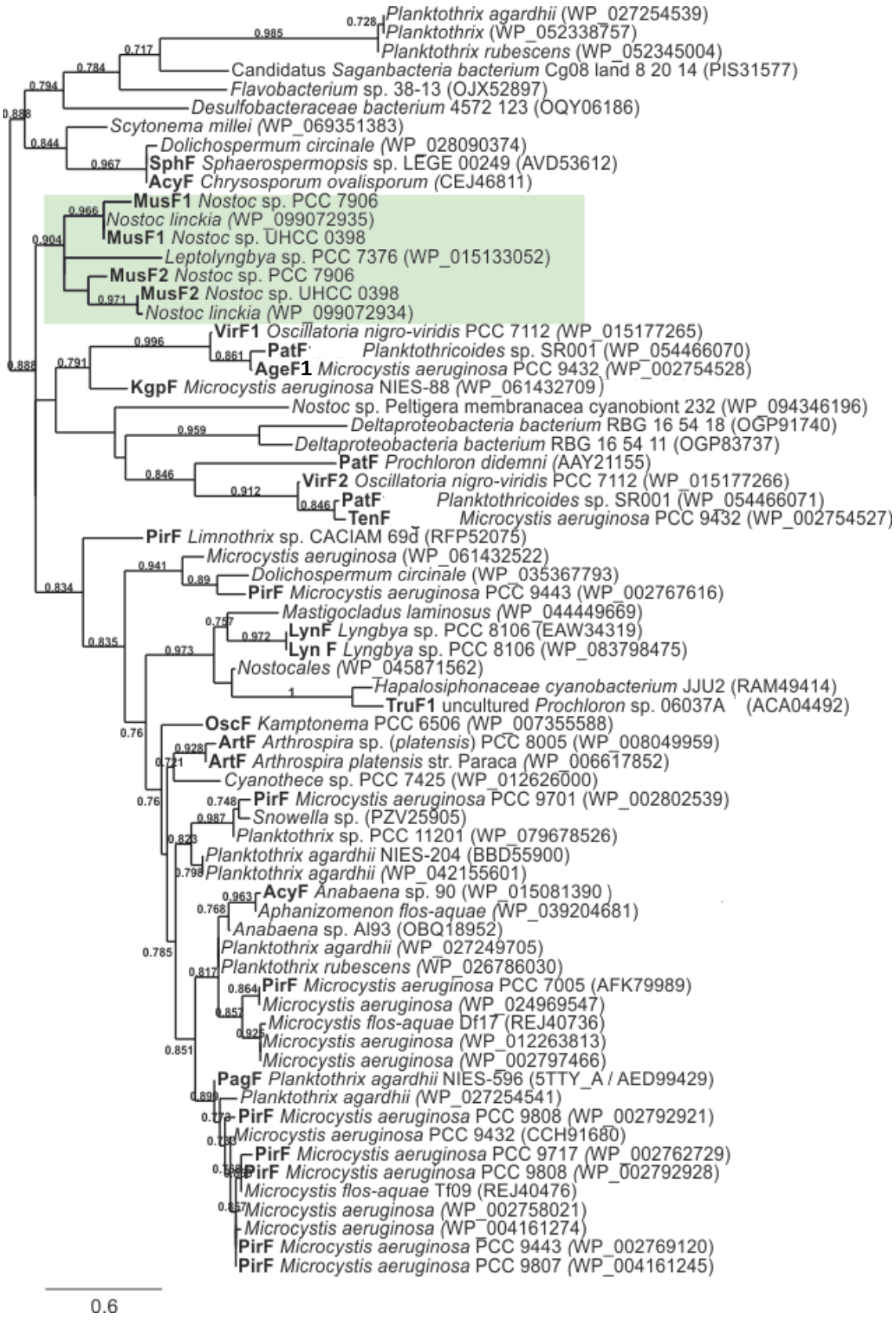
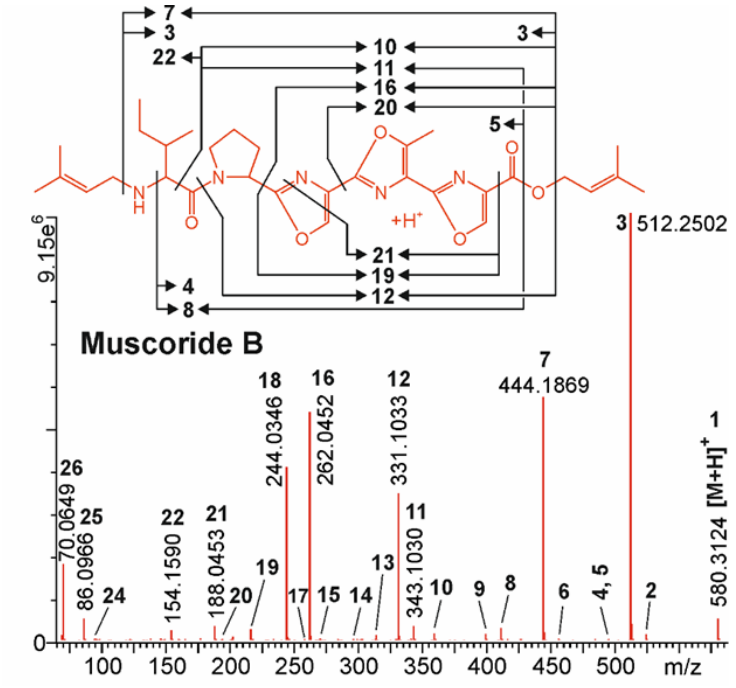


Figure 6. Phylogenetic tree of cyanobactin prenyltransferases. The muscoride prenyltransferases, MusF1 and MusF2, are highlighted in light grey. SH-like values above 0.7 are shown at the node.

5.3 Identification and purification of muscoride B from *Nostoc* sp. UHCC 0398

Inspection of the MusE precursor peptide core sequence of the *Nostoc* sp. UHCC 0398 *mus* biosynthetic gene cluster indicated that the muscoride chemical variant encoded by this strain would have three contiguous oxazoles (Figure 4). The mass spectrometry analysis by UPLC-HDMS of the methanol crude extract from *Nostoc* sp. UHCC 0398 confirmed the production of the predicted bis-prenylated pentapeptide muscoride variant (*m/z* 580.3124, Δ-1.0 ppm), here on named muscoride B (Figure 7).



	Ion assignment		B
No	Ion assignment without NL	Neutral losses (NL)	*m/z*
2	$[M+H]^+$	C_4H_8	524.2493
6	$[M+H]^+$	Pr, C_4H_8	456.1873
13	$[M+H]^+$	Ile/Val, 2xPr, NH_3	314.0762
14	Pro(Ozn)-mOzn-OznA-Pr/Pro(mOzn)-mOznA-Pr $+H^+$	Pr-OH, NH_3	296.0652
15	Pro(Ozn)-mOzn-Ozn/Pro(mOzn)-mOzn $+H^+$	NH_3	270.0860
17	$[M+H]^+$	Pr-Ile, Pr, CO_2, CH_3N	258.0867
18	Ozn-mOzn-OznA/mOzn-mOznA $+H^+$	H_2O	244.0346
21	Ozn-mOzn-Ozn/mOzn-mOzn $+H^+$	CH_2O	188.0453
23	Ozn-mOzn/mOzn-mOzn $+H^+$	C_2H_4O	107.0252
24	Ozn-CN/mOzn-CN $+H^+$	-	95.0236
25	Ile/Val immonium	-	86.0966
26	Pro immonium	-	70.0649

Figure 7. Annotated UPLC-HDMS product ion mass spectra (MS^2) of protonated $[M+H]^+$ muscoride B with the structure and fragmentation schemes (provided courtesy of Dr. Jouni Jokela).

Muscoride B was eluted from the SPE cartridges with methanol (Figure 8; Figure S1 and Figure S2 in Appendix 2). The muscoride B fractions were further purified twice using HPLC-MS (Figure 9, Figure 10). The first round did not purify the muscoride B enough but after the second purification with HPLC-MS muscoride B was sufficiently pure for NMR spectroscopy (Figure 9, Figure 10). According to the NMR spectroscopy the yield of muscoride B was 20 µg.

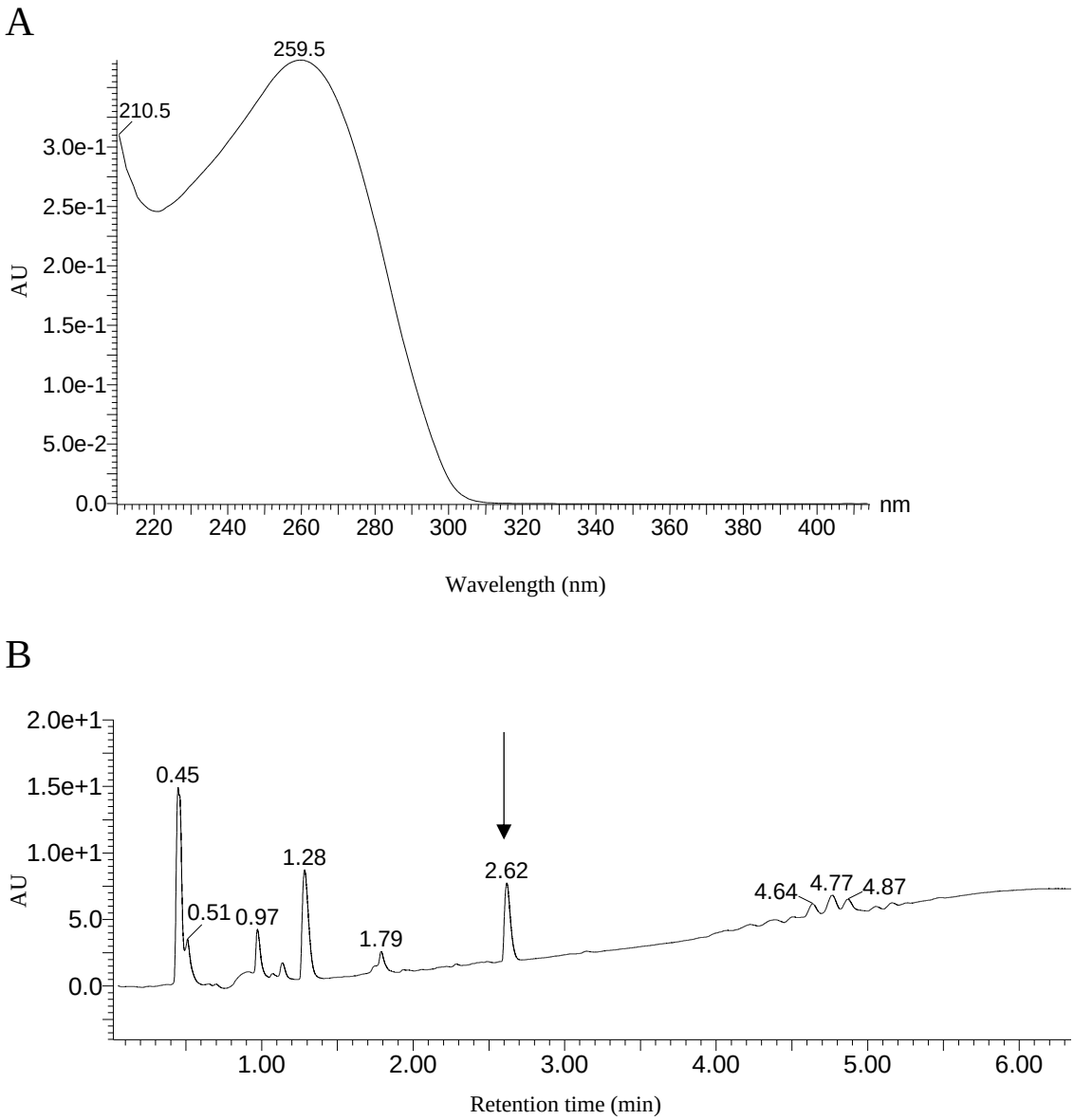


Figure 8. In panel A the UV absorbance spectra of muscoride B with maxima peak at 259.5 nm. In panel B the corresponding peak for muscoride B at 2.62 min in a UPLC-HDMS chromatogram from a SPE methanol extract (provided courtesy of Mr. Matti Wahlsten).

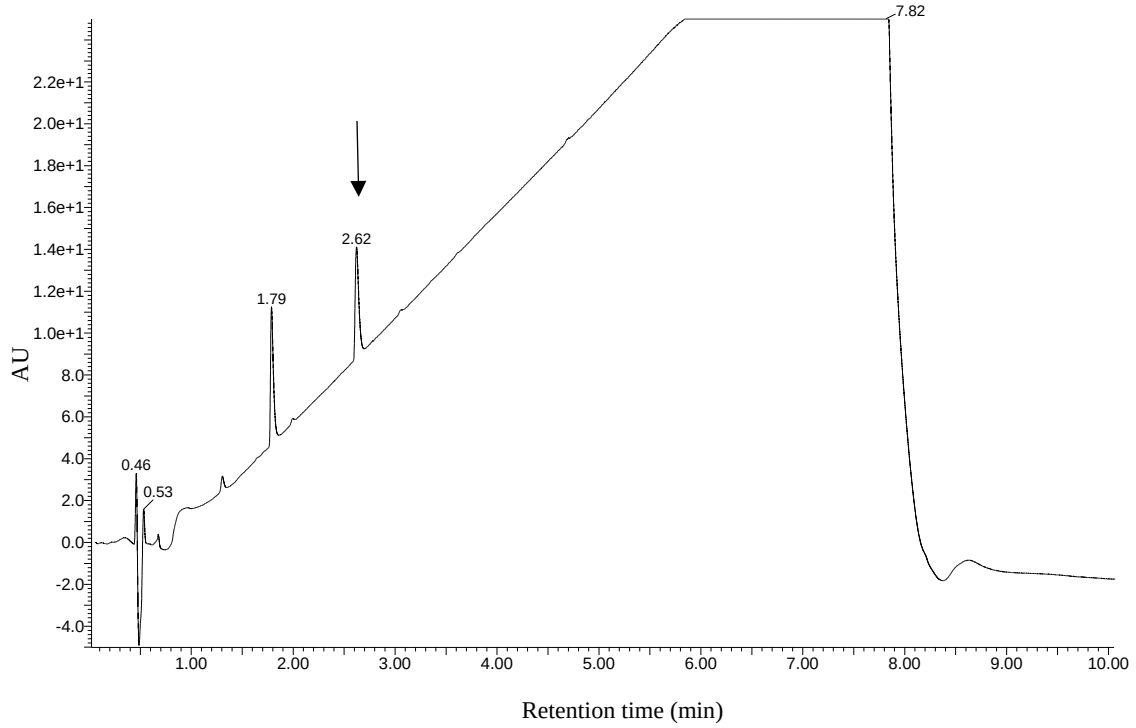


Figure 9. UPLC-HDMS chromatogram after the first HPLC-MS purification shows muscoride B peak at 2.62 min but also impurities (provided courtesy of Mr. Matti Wahlsten).

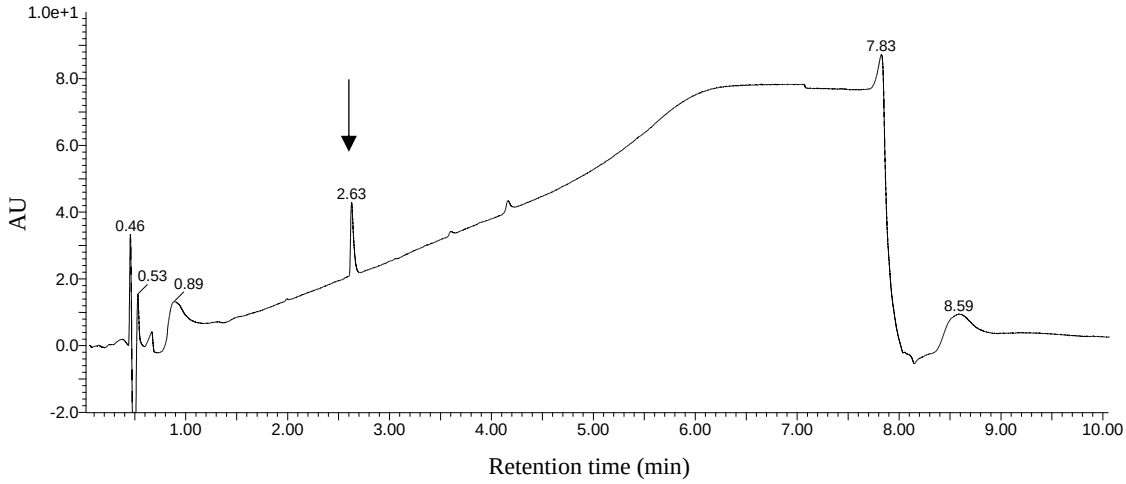


Figure 10. UPLC-HDMS chromatogram after the second HPLC-MS purification shows muscoride B peak at 2.63 min (provided courtesy of Mr. Matti Wahlsten).

5.4 Structural analysis of muscoride B

Comparing proton and carbon chemical shift (δ_H and δ_C) values obtained from the one-dimensional (1D) NMR analysis of muscoride B with values from the literature, shows that muscoride B is forward prenylated at the N- and C-terminus (Table 4, Figure 11). The N-terminal δ_H and δ_C values from aeruginosamide A and virenamide A, and the C-terminal values from muscoride A correspond to the N- and C-terminal values from muscoride B (Table 4). Different two-dimensional (2D) NMR spectroscopy experiments ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HMBC and ROESY show correlations in muscoride B (Figure 11). The NMR spectroscopy spectra are presented in Figures S1-S5 in Appendix 3 and the collected spectral data in Table S1 in Appendix 4.

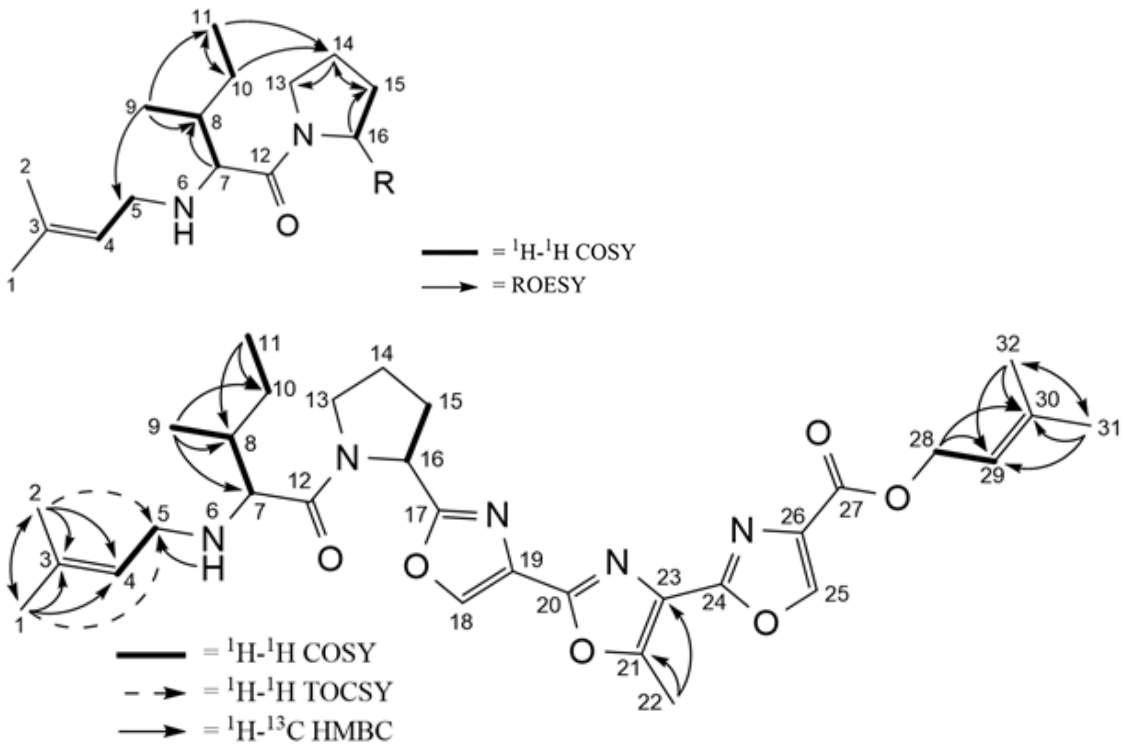


Figure 11. In upper structure ROESY (arrows) and ^{1}H-^{1}H COSY (bold lines) and in lower structure ^{1}H-^{1}H COSY (bold lines), ^{1}H-^{1}H TOCSY (cut arrows) and ^{1}H-^{13}C HMBC (arrows) showing correlations in the muscoride B structure (provided courtesy of Dr. Jouni Jokela).

Table 4. The carbon and proton chemical shift values of muscoride B, muscoride A, aeruginosamide A and virenamide A forward prenyl groups.

	Muscoride B forward prenyl in pyridine-d_5. Minor rotamer in parenthesis			Muscoride A forward prenyl in $CDCl_3$ (Nagatsu et al. 1995)		Aeruginosamide A forward prenyl in $CDCl_3$ (Lawton et al. 1999)		Virenamide A forward prenyl in $CDCl_3$ (Carroll et al. 1996)	
Location	C-atom no	δ_C (ppm)	δ_H (ppm)	δ_C (ppm)	δ_H (ppm)	δ_C (ppm)	δ_H (ppm)	δ_C (ppm)	δ_H (ppm)
N-terminal	1	25.6	1.68 (1.54)			26.1	1.66	25.7	1.69
	2	17.9	1.62 (1.49)			18.2	1.55	18.0	1.54
	3	133.6	-			134.6		135.4	-
	4	124.6	5.43			122.5	5.12	121.9	5.13
	5	49.4	3.38			48.0	3.02	48.1	3.08
	5’		3.24				3.10		
C-terminal	28	nd	4.96	61.8	4.82				
	29	119.0	5.53	118.6	5.46				
	30	139.3	-	139.1					
	31	25.3	1.66	25.8	1.77				
	32	17.7	1.67	18.1	1.76				

nd=not detected

5.5 Prenylation assay

A prenylation assay was performed using synthetic amino and carboxyl substrates with purified MusF1 and MusF2 prenyltransferase to test the regiospecificity of the prenyltransferases (Figure 12, Figure 13). The prenylation for both amino and carboxyl substrates were detected in all reaction mixtures (Figure 12, Figure 13, Table S1 in Appendix 5). The carboxyl substrate was prenylated by the purified MusF1 prenyltransferase (Figure 12). The prenylation reaction for the amino substrate with the purified MusF2 prenyltransferase produced a major (*m/z* 477) and minor (*m/z* 140) prenylated product (Figure 13). Slightly more products were produced in reaction mixture 2. and slightly in the lower temperature (28°C), except for the minor amino product (Table S1 in Appendix 5).

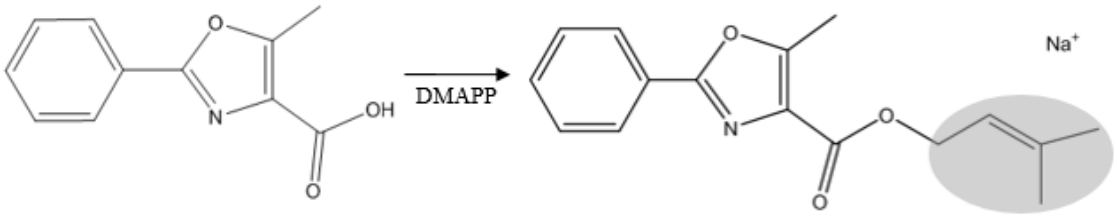


Figure 12. Chemical structure of the carboxyl substrate (5-methyl-2-phenyl-1,3-oxazole-4-carboxylic acid) before and after the prenyl group (highlighted in gray) is added (provided courtesy of Mr. Matti Wahlsten).

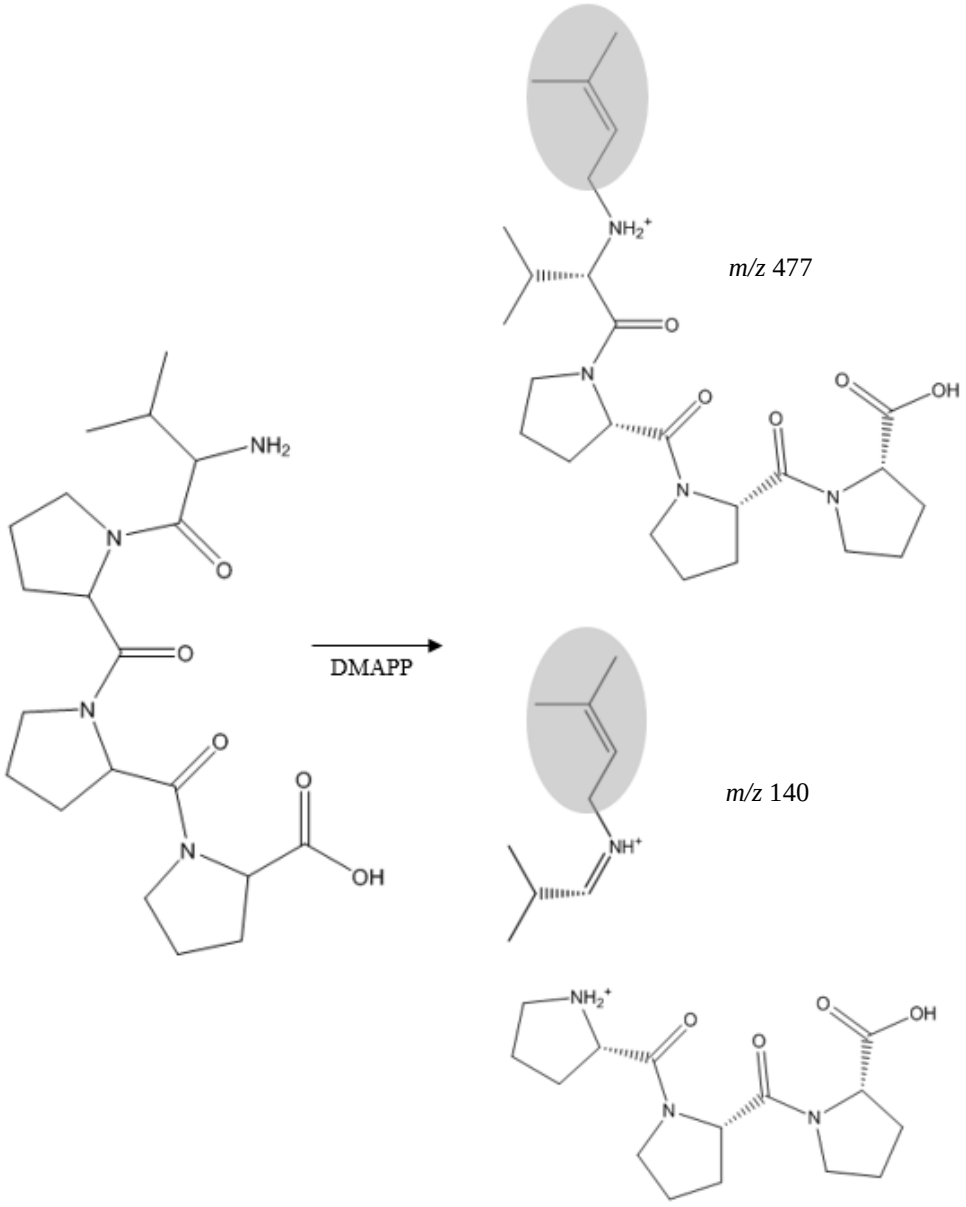


Figure 13. Chemical structure of the amino substrate (VPPP) before and after the major (*m/z* 477) and minor (*m/z* 140) products formed after the prenylation. Prenyl group in highlighted in gray (provided courtesy of Mr. Matti Wahlsten).

5.6 Antimicrobial assay

The semi-pure extract of muscoride B from *Nostoc* sp. UHCC 0398 showed antimicrobial activity against *Bacillus cereus* (1881) with an inhibition zone 10 mm (diameter) in the agar well diffusion method assay (Figure 14). The inhibition zone for the positive control was 30 mm (diameter) and no inhibition zone for the negative control (Figure 14).

No activity was detected in the disk diffusion assay for the semi-pure muscoride B extract from *Nostoc* sp. UHCC 0398 or in the agar well diffusion assay for the methanol crude extracts from *Nostoc* sp. UHCC 0398 and *Nostoc* sp. PCC 7906.

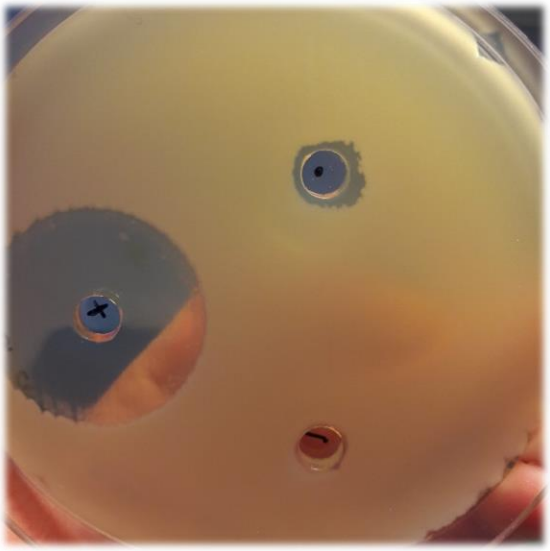


Figure 14. Agar well diffusion method with semi-pure muscoride B from *Nostoc* sp. UHCC 0398 showing antimicrobial activity against *B. cereus*.

6 DISCUSSION

6.1 The muscoride biosynthetic pathway

This study confirmed that muscoride A is synthesized by the cyanobactin biosynthetic pathway and the muscoride (*mus*) biosynthetic pathway was identified. Cyanobactin biosynthetic pathways consists of typical modification enzymes and a precursor peptide with conserved recognition sequences and at least one core sequence (Arnison et al. 2013). A cyanobactin biosynthetic gene cluster, with several of the these cyanobactin biosynthetic genes and a core sequence that encoded the amino acids for muscoride A, was identified from the genome of *Nostoc* sp. PCC 7906. This discovery indicated that muscoride A is synthesized by the cyanobactin biosynthetic pathway as hypothesized (Leikoski et al. 2013) and belongs to the cyanobactin family.

Based on the known chemical structure of muscoride A, the discovered core peptide sequence and 10 biosynthetic genes, a putative biosynthetic scheme for the biosynthesis of muscoride A was suggested (Figure 5). The precursor peptide MusE had a single cassette encoding one core sequences with four amino acids, Val, Pro, Thr, Thr (Figure 4). The *mus* pathway encoded a adenylate cyclase-like MusD heterocyclase, which indicated that the first modification of the core peptide is the formation of methyloxazolines from the Thr encoded in the core sequence (McIntosh & Schmidt 2010, McIntosh et al. 2010a). This reaction is known to be directed by a conserved recognition sequence, RSI, which in the *mus* biosynthetic pathway was QLDLSEEEL (Sardar et al. 2015). Methyloxazolines can further be oxidized into methyloxazoles by a dehydrogenase (Donia & Schmidt 2011). The mature muscoride A has two methyloxazoles (Nagatsu et al. 1995) that are formed when the methyloxazolines are oxidized by the MusOx dehydrogenase (Figure 5).

The following modification in cyanobactin biosynthetic pathways is the cleavage of the core peptide from the precursor peptide by two proteases A and G (Lee et al. 2009). Two conserved recognition sequences, RSII and RSIII, located in front of and after the core sequence directs these reactions (Sardar et al. 2015). In the *mus* biosynthetic pathway RSII (GVSPS) directs the MusA protease to liberate N-termini of the core peptide (Figure 5). Surprisingly, the RSIII and the G protease and the macrocyclase is missing from the muscoride A gene cluster (Figure 5). The precursor peptide has two amino acids, Gly and Val flanking the core peptide at the C-termini, that somehow are cut out from the mature

peptide (Figure 5). The biosynthetic pathways for the linear cyanobactins aeruginosamide B, C and viridisamide A all encode the G protease and the RSIII (Leikoski et al 2013). In 2017 Sardar et al. (2017) showed that the G protease is responsible for the linearization of the C-termini in the aeruginosamide (B) pathway instead of the macrocyclization. How the two amino acids are removed from the C-termini of the precursor peptide MusE, remains unclear.

Methyl- and prenyltransferases are often encoded in cyanobactin biosynthetic gene clusters, but can sometimes be inactive (Sivonen et al. 2010, Leikoski et al. 2013). Viridisamide A, aeruginosamide A, B and C have N-prenylated and C-methylated termini (Lawton et al. 1999, Leikoski et al 2013). In the aerugionsamide pathway a didomain methyl/prenyltransferase is responsible for the modifications of the termini (Sardar et al. 2017). Muscoride A is not methylated, but instead prenylated at the N- and C-termini (Nagatsu et al. 1995). The *mus* gene cluster encoded two prenyltransferases, MusF1 and MusF2 (Figure 5). It was not clear from the bioinformatic data which prenyltransferase makes the modifications. The prenylation assay suggested that the prenyltransferase are regiospecific whereas MusF1 would be responsible for the prenylation of the carboxyl termini and MusF2 for the prenylation of the amino termini in the *mus* biosynthetic pathway (Figure 12, Figure 13). Hence, the final modification of the core peptide was the prenylation by prenyltransferase MusF1 to yield a forward prenyl group at the carboxyl-terminus and MusF2 to yield a reverse prenyl group at the amino-terminus of muscoride A (Figure 5).

The biosynthetic pathways of aeruginosamide, viridisamide, patellin/trunkamide all encode two prenyltransferase, but no know bisprenylated cyanobactins have been described before (Donia et al. 2008, Sivonen et al. 2010, Leikoski et al 2012, 2013). The phylogenetic analysis of prenyltransferase enzymes showed that MusF1 and MusF2 were distantly related to VirF1 and AgeF1, which catalyze the N-prenylation of linear cyanobactins, from the viridisamide and aeruginosamide biosynthetic gene clusters (Leikoski et al 2013) (Figure 6). The MusF1 and MusF2 enzymes fell into two distinct sister clades suggesting they arose through gene duplication (Figure 6).

6.2 The detection and characterization of a novel muscoride variant

A new muscoride variant, muscoride B was identified and characterized form *Nostoc* sp. UHCC 0398 by genome mining (Figure 11). The chemical structure of this polyoxazole

bisprenylated linear pentapeptide was determined by bioinformatic predictions and analytical chemistry (Figure 11). After genome sequencing has become more efficient and the collected data easily available in public databases the bioinformatic approach for discovery of novel natural products has increased in the past decade (Albarano et al. 2020). Using bioinformatic tools, the known bioinformatic data can be used for the identification of novel biosynthetic pathways and compounds and many cyanobactin gene clusters and cyanobactins have been detected this way (Dittmann et al. 2015, Albarano et al. 2020). Here, the identification of the cyanobactin biosynthetic pathway from *Nostoc* sp. PCC 7906 lead to the discovery of *mus* gene cluster that could be identified from 12 *Nostoc* sp. strains through bioinformatic analysis (Figure 4). The *mus* gene cluster detected from *Nostoc* sp. UHCC 0398, encoded a core sequence that by bioinformatic predictions and chemical analysis suggested to be a linear pentapeptide polyoxazole with bisprenylated termini (Figure 4). By these findings it was not yet to possible to verify the structure of this novel peptide, so the chemical structure was further analyzed by NMR spectroscopy.

Characterization of natural products by NMR spectroscopy is a frequently used method which has improved during the last decades (Breton and Reynolds 2013). By NMR spectroscopy it is possible to determine the molecular structure and purity of compounds from samples less than 1 mg (Breton and Reynolds 2013). Comparing the obtained NMR spectroscopy data from muscoride B with literature describing other linear natural products from cyanobacteria, confirmed the structure of muscoride B and the forward orientation of the prenyl groups at the N- and C-terminus (Figure 11). The carbon and proton chemical shift values from forward prenylated muscoride A, aeruginosamide A and virenamide A matched the values for muscoride B (Nagatsu et al. 1995, Carroll et al. 1996, Lawton et al. 1999). Based on the results from the bioinformatic and chemical analysis the structure of muscoride B was determined (Figure 11).

Only five linear cyanobactins have been reported earlier (Leikoski et al. 2013, Crnkovic et al. 2020). The identification of the cyanobactin biosynthetic gene cluster *mus* from *Nostoc* sp. PCC 7906 and *Nostoc* sp. UHCC 0398 and the characterization of the muscoride variant reviles two new linear cyanobactins, muscoride A and muscoride B (Figure 4, Figure 11). The chemical structure of muscoride A and B are unique since no polyoxazole containing cyanobactins have been described before (Sivonen et al. 2010). In RiPPs contiguous (methyl)oxazoles has been reported from the linear azol(in)e-containing peptide plantazolicin A and B produced by *Bacillus amyloliquefaciens* FZB42

(Kalyon et al. 2011, Arnison et al. 2013). Other natural product with contiguous oxazoles are linear peptide mechercharmycin B, cyclic peptides mechercharmycin A, YM-216391, urukthapelstatin A and in the bicyclic peptides diazonamides A-E (Lindquist et al. 1991, Kanoh et al. 2005, Sohda et al. 2005, Matsuo et al. 2007, Fernández et al. 2008).

Though linear cyanobactins and polyoxazoles are rare in the nature, putative and novel compounds are described constantly. In 2017, Te et al. reported a putative cyanobactin gene cluster from *Planktothriocoides* sp. SR001, a free-living filamentous cyanobacterium from the order Oscillatoriales (Te et al. 2017). It had eight biosynthetic genes with similarity to the viridisamid A and aeruginosamide and a precursor peptide with conserved leader sequence with the conserved motif LAELSEE and a multiple core encoding three repeats of the core sequence FVPCC (Leikoski et al. 2013, Te et al. 2017). The chemical structure of the putative cyanobactin was not described, but the putative cyanobactin biosynthetic pathway could produce a linear cyanobactin (Te et al. 2017). In 2013, approximately 200 cyanobactins were characterized, mostly by traditional methods whereas today 513 experimentally characterized RiPPs are listed in the RiPPMiner (Arnison et al. 2013, Agrawal et al. 2017).

6.3 Purification of muscoride B

The purification of the muscoride variant, muscoride B from *Nostoc* sp. UHCC 0398 yield 20 µg. The purification of muscoride B was highly challenging and several additional steps were added to the first tested purification schema before a pure enough muscoride B was obtained for NMR spectroscopy.

The first tested method was based on the purification of muscoride A and followed commonly used methods for natural product isolation, using solvent extraction and isolation by chromatography (Nagatsu et al. 1995, Zhang et al. 2018). During the extraction and separation, a white greasy precipitation was formed. The precipitation was analyzed with UPLC-HDMS, but the substance could not be determined. Also, the possibility that muscoride B would have been transfer into the precipitation was tested, but luckily was not detected from the precipitation. Impurities were also detected after muscoride B had been fractionated with HPLC-MS (Figure 9).

To optimization of the purification method, a pre-test to analyze the yield after the extraction was made in a smaller scale, different cultivation batches of *Nostoc* sp. UHCC 0398 were tested and the extraction and separation solvents analyzed separately. None of

the previous mentioned tests showed anything abnormal. Two different columns, different settings and solvents were also tested for the HPLC-MS.

The purification method was improved by a few modifications and additional steps, and the final method that led to the purification of muscoride B is described in the thesis. Glass wear were used during the purification as far as possible to avoid any contamination from plastic. After the extraction of *Nostoc* sp. UHCC 0398 freeze-dried cells, the compounds in the methanol extract were bound to octadecylsilyl silica gel to minimize the compounds attaching to the rotating flask during the rotary evaporation. Though the muscoride B sample had to fractionated twice with HPLC-MS due to impurities, no white precipitation was formed this time (Figure 9, Figure 10).

The yield of muscoride B from the purification was little, but fortunately enough for structural analysis by NMR spectroscopy. The yield might have been affected by the numerous steps in the purification, by some loss of compound. Also, the cultivation conditions can affect the productivity of natural products and the amount of synthesized metabolite can vary between cultivation batches and in some cases, the metabolite might not be produced at all in laboratory conditions (Pettit 2009).

6.4 Antibiotic activity of muscoride B

According to these preliminary results, a semi-pure extract of muscoride B showed antimicrobial activity against *B. cereus* (Figure 14). Due to the low purification yield of muscoride B, no analysis could be performed with the pure compound. Using the semi-pure compound, it was difficult to know how much of muscoride B was present in the sample. Results showed in the agar well diffusion method with a more concentrated sample than in the disk diffusion method assay (Figure 14). A more concentrated semi-pure extract might have shown antimicrobial activity in the disk diffusion method assay as well. Further analysis with pure muscoride B should be done to verify the antimicrobial activity and possible minimum inhibitory concentration values for both muscoride A and B, which could start with analyzing different amounts of methanol crude extracts from freeze dried cells of *Nostoc* sp. UHCC 0398 and *Nostoc* sp. PCC 7906.

Many cyanobactins are reported to have antimicrobial activity including muscoride A (Nagatsu et al. 1995, Sivonen et al. 2010). Muscoride A has antimicrobial activity against *Bacillus subtilis*, but no further bioactive studies have been done after 1995 (Nagatsu et al. 1995). Based on the chemical structures of muscoride A and B, possible

mechanisms of action for the biological activity could lie in the oxazoles or the prenyl groups (Botta et al. 2005, Mhlongo et al. 2020). Other natural products with oxazoles have shown antibacterial, antiviral, anti-algal and cytotoxic activity (Mhlongo et al. 2020). Linear RiPPs with oxazoles such as plantazoline A has antibiotic activity against *Bacillus anthracis* and microcin B17 produced by *Escherichia coli* is a DNA-gyrase inhibitor (Yorgey et al. 1994, Molohon et al. 2011). Cytotoxic activity has been reported from cyclic peptides mechercharmycin A from *Thermoactinomyces* sp. YM3-251, YM-216391 from *Streptomyces noblilis* and urukthapelstatin A from *Mechercharimyces asporophorigenens* YM11-542 with contiguous oxazoles (Kanoh et al. 2005, Sohda et al. 2005, Matsuo et al. 2007). The chemical structure of YM-216391 has similarity to the potent telomerase inhibitor telomestatin (Sohda et al. 2005, Shin-ya et al. 2001). Also, the bicyclic peptides diazonamides A-E, with contiguous oxazoles, from the ascidian *Diazona angulata* have displayed cytotoxic activity (Lindquist et al. 1991, Fernández et al. 2008). The linear cyanobactins scytodecamide and aeruginosamide A have shown cytotoxic activity whereas the other linear cyanobactins viridisamide A, aeruginosamide B and C have not been evaluated for biological activity (Lawton et al. 1999, Leikoski et al. 2013, Crnkovic et al. 2020). In addition to further antimicrobial testing of muscoride A and B, the cytotoxicity could also be evaluated.

7 CONCLUSIONS

This study confirmed the biosynthetic origin of muscoride and expanded the cyanobactin family with two linear bisprenylated polyoxazole peptides. The identification of a cyanobactin biosynthetic gene cluster from *Nostoc* sp. PCC 7906, showed that muscoride A belong to the cyanobactin family and the muscoride (*mus*) biosynthetic pathway was described. In addition to characteristic cyanobactin biosynthetic genes and post-translational modification, interesting features were recognized from the *mus* biosynthetic pathway. Two regiospecific prenyltransferases were described for the first time from a cyanobactin biosynthetic pathway and key elements needed for the liberation of the C-terminus were missing. Further bioinformatic analysis revealed several *Nostoc* strains encoding the characterized *mus* gene cluster and lead to the discovery of a novel antimicrobial muscoride variant, muscoride B. This linear bis-prenylated pentapeptide had a rare polyoxazole chemical structure and showed antimicrobial activity. Muscoride A and muscoride B have rare chemical structures that might possess undiscovered bioactive properties and biosynthetic genes with unusual mechanism of action which leaves room for further research. These findings furthermore show the great variety among cyanobactins and that the cyanobactin family is still growing. This study also showed the power of genome mining, that serves as an effective tool for the discovery of novel compounds, and that cyanobacteria keep on providing natural products with unique chemical structures and potential compounds or leads for the pharmacological industry.

ACKNOWLEDGMENTS

First, I would like to express my deepest gratitude to my supervisor Docent David Fewer for the guidance and support throughout my thesis. I would like to thank doctoral student Antti Mattila for all the advices and motivational conversations. I wish to say my grateful thanks to Dr. Jouni Jokela and Mr. Matti Wahlsten for teaching me about analytical chemistry and for the technical assistance and to M.Sc. Lyudmila Saari for familiarizing me with cultivation of cyanobacteria. I would also like to thank all the other members in the cyanobacteria research group - it was a pleasure to work in a such a helpful, warm and welcoming group. Finally, I would like to thank my dear family and friends for the endless support, with a special thanks to my beloved mother Päivi for inspiration and being there.

REFERENCES

Agrawal P, Khater S, Gupta M, Sain N & Mohanty D. 2017. RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. Nucleic Acids Research 45, W1: W80–W88, doi: 10.1093/nar/gkx408.

Agarwal V, Pierce E, McIntosh J, Schmidt EW & Nair SK. 2012. Structures of cyanobactin maturation enzymes define a family of transamidating proteases. Chemistry and Biology 19, 11: 1411-1422.

Albarano L, Esposito R, Ruocco N & Constantin M. 2020. Genome mining as new challenge in natural products discovery. Marine Drugs 18, 4, article number 18040199.

Amaike K, Muto K, Yamaguchi J & Itami K. 2012. Decarbonylative C-H coupling of azoles and aryl esters: Unprecedented nickel catalysis and application to the synthesis of muscoride A. Journal of the American Chemical Society 134, 33: 13573-13576.

Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian K-D, Fischbach MA, Garavelli JS, Göransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Müller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJT, Rebuffat S, Ross RP, Sahl H-G, Schmidt EW, Selsted ME, Severinov K, Shen B, Sivonen K, Smith L, Stein T, Süssmuth RD, Tagg JR, Tang G-L, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM & Van Der Donk WA. 2013. Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. Natural Product Reports 30, 1: 108-160.

Bagley MC, Buck RT, Hind SL & Moody CJ. 1998. Synthesis of functionalised oxazoles and bis-oxazoles. Journal of the Chemical Society - Perkin Transactions 1, 3: 591-600.

Bauer AW, Kirby WMM, Sherris JC & Turk M. 1966. Antibiotic susceptibility testing by a standardized single disk method. American Journal of Clinical Pathology 45, 4: 493-496.

Bent AF, Koehnke J, Houssen WE, Smith MCM, Jaspars M & Naismith JH. 2013. Structure of PatF from *Prochloron didemni*. Acta Crystallographica Section F: Structural Biology and Crystallization Communications 69, 6: 618-623.

Bent AF, Mann G, Houssen WE, Mykhaylyk V, Duman R, Thomas L, Jaspars M, Wagner A & Naismith JH. 2016. Structure of the cyanobactin oxidase ThcOx from *Cyanothece* sp. PCC 7425, the first structure to be solved at Diamond Light Source beamline I23 by means of S-SAD. Acta Crystallographica Section D: Structural Biology 72, 11: 1174-1180.

Botta B, Vitali A, Menendez P, Misiti D & Monache GD. 2005. Prenylated flavonoids: Pharmacology and biotechnology. Current Medicinal Chemistry 12, 6: 713−739.

Breton RC & Reynolds WF. 2013. Using NMR to identify and characterize natural products. Natural Product Reports 30, 4: 501-524.

Burja AM, Banaigs B, Abou-Mansour E, Grant Burgess J & Wright PC. 2001. Marine cyanobacteria - A prolific source of natural products. Tetrahedron 57, 46: 9347-9377.

Burkhart BJ, Hudson GA, Dunbar KL & Mitchell DA. 2015. A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. Nature Chemical Biology 11, 8: 564-570.

Carmichael WW. 1992. Cyanobacteria secondary metabolites—the cyanotoxins. Journal of Applied Bacteriology 72, 6: 445-459.

Carroll AR, Feng Y, Bowden BF & Coll JC. 1996. Studies of Australian ascidians. 5. Virenamides A-C, new cytotoxic linear peptides from the colonial Didemnid ascidian *Diplosoma virens*. Journal of Organic Chemistry 61, 12: 4059-4061.

Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG & Parkhill. 2005. ACT: the Artemis Comparison Tool. Bioinformatics 21, 16: 3422-3423.

Castenholz RW. 2015. General Characteristics of the Cyanobacteria. In e-book: Whitman WB (ed). Bergey´s manual of systematics of archaea and bacteria. John Wiley & Sons, Inc.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17, 540-552.

Cohen Y, Jørgensen BB, Revsbech NP & Poplawski R. 1986. Adaptation to hydrogen sulfide of oxygenic and anoxygenic photosynthesis among cyanobacteria. Applied Environmental Microbiology 51, 2: 398–407.

Coqueron PY, Didier C & Ciufolini MA. 2003. Iterative oxazole assembly via α-chloroglycinates: Total synthesis of (-)-muscoride A. Angewandte Chemie - International Edition 42, 12: 1411-1414.

Correa A, Cornella J & Martin R. 2013. Nickel-catalyzed decarbonylative C-H coupling reactions: A strategy for preparing Bis(heteroaryl) backbones. Angewandte Chemie - International Edition 52, 7: 1878-1880.

Crnkovic CM, Braesel J, Krunic A, Eustáquio AS & Orjala J. 2020. Scytodecamide from the cultured *Scytonema* sp. UIC 10036 expands the chemical and genetic diversity of cyanobactins. ChemBioChem 21, 6: 845-852.

Dereeper A, Guignon, V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM & Gascuel O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research 36 (Web Server issue), W465-W469. https://doi.org/10.1093/nar/gkn180

Dittmann E, Gugger M, Sivonen K & Fewer DP. 2015. Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. Trends in Microbiology 23, 10: 642-652.

Dittmann E, Neilan BA, Erhard M, Von Döhren H & Börner T. 1997. Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. Molecular Microbiology 26, 4: 779-787.

Donia MS, Hathaway BJ, Sudek S, Haygood MG, Rosovitz MJ, Ravel J & Schmidt EW. 2006. Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. Nature Chemical Biology 2, 12: 729-735.

Donia MS, Ravel J & Schmidt EW. 2008. A global assembly line for cyanobactins. Nature Chemical Biology 4, 6: 341-343.

Donia MS and Schmidt EW. 2010. Cyanobactins – Ubiquitous cyanobacterial ribosomal peptide metabolites. In: Mander L and Lui H-W (eds.). Comprehensive Natural Products II Chemistry and Biology. Oxford: Elsviser. 2, pp. 539-558.

Donia MS & Schmidt EW. 2011. Linking chemistry and genetics in the growing cyanobactin natural products family. Chemistry and Biology 18, 4: 508-519.

Dunbar KL, Chekan JR, Cox CL, Burkhart BJ, Nair SK & Mitchell DA. 2014. Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. Nature Chemical Biology 10, 10: 823-829.

Dunbar KL, Melby JO & Mitchell DA. 2012. YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. Nature Chemical Biology 8, 6: 569-575.

Edgar, RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 5: 1792-97.

Fernández R, Martin MJ, Rodríguez-Acebes R, Reyes F, Francesch A & Cuevas C. 2008. Diazonamides C–E, new cytotoxic metabolites from the ascidian *Diazona* sp. Tetrahedron Letters 49, 14: 2283–2285.

Funk MA & van der Donk WA. 2017. Ribosomal natural products, tailored to fit. Accounts of Chemical Research 50, 7: 1577-1586.

Gao S, Ge Y, Bent AF, Schwarz-Linek U & Naismith JH. 2018. Oxidation of the cyanobactin precursor peptide is independent of the leader peptide and operates in a defined order. Biochemistry 57, 41: 5996-6002.

Gu W, Dong S-H, Sarkar S, Nair SK & Schmidt EW. 2018a. The biochemistry and structural biology of cyanobactin pathways: Enabling combinatorial biosynthesis. Methods in Enzymology 604: 113-163.

Gu W, Sardar D, Pierce E & Schmidt EW. 2018b. Roads to Rome: Role of multiple cassettes in cyanobactin RiPP biosynthesis. Journal of the American Chemical Society 140, 47: 16213-16221.

Gu W & Schmidt EW. 2017. Three principles of diversity-generating biosynthesis. Accounts of Chemical Research 50, 10: 2569-2576.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W & Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology 59, 3: 307-321.

Hao Y, Pierce E, Roe D, Morita M, McIntosh JA, Agarwal V, Cheatham III TE, Schmidt EW & Nair SK. 2016. Molecular basis for the broad substrate selectivity of a peptide prenyltransferase. Proceedings of the National Academy of Sciences of the United States of America 113, 49: 14037-14042.

Ireland C & Scheuer PJ. 1980. Ulicyclamide and Ulithiacyclámide, Two new small peptides from a marine tunicate. Journal of the American Chemical Society 102, 17: 5688-5691.

Kanoh K, Matsuo Y, Adachi K, Imagawa H, Nishizama M & Shizuri Y. 2005. Mechercharmycins A and B, cytotoxic substances from marine-

derived *Thermoactinomyces* sp. YM3-251. The Journal of Antibiotics 58, 4: 289–292.

Katz L & Baltz RH. 2016. Natural product discovery: past, present, and future. Journal of Industrial Microbiology and Biotechnology 43, 2-3: 155-176.

Kalyon B, Helaly SE, Scholz R, Nachtigall J, Vater J, Borriss R & Süssmuth RD. 2011. Plantazolicin A and B: Structure elucidation of ribosomally synthesized thiazole/oxazole peptides from *Bacillus amyloliquefaciens* FZB42. Organic Letters 13, 12: 2996− 2999.

Knoll AH. 2008. Cyanobacteria and earth history. In: Herrero A & Flores E (eds.). The cyanobacteria – molecular biology, genomics and evolution. Norfolk: Caister Academic Press. pp. 1–19.

Koehnke J, Bent A, Houssen WE, Zollman D, Morawitz F, Shirran S, Vendome J, Nneoyiegbe AF, Trembleau L, Botting CH, Smith MCM, Jaspars M & Naismith JH. 2012. The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. Nature Structural and Molecular Biology 19, 8: 767-772.

Koehnke J, Mann G, Bent AF, Ludewig H, Shirran S, Botting C, Lebl T, Houssen WE, Jaspars M & Naismith JH. 2015. Structural analysis of leader peptide binding enables leader-free cyanobactin processing. Nature Chemical Biology 11, 8: 558-563.

Kótai J. 1972. Instructions for preparation of modified nutrient solution Z8 for algae. Norwegian Institute for Water Research, Oslo.

Lawton LA, Morris LA & Jaspars M. 1999. A bioactive modified peptide, aeruginosamide, isolated from the cyanobacterium *Microcystis aeruginosa*. Journal of Organic Chemistry 64, 14: 5329-5332.

Lee J, Mcintosh J, Hathaway BJ & Schmidt EW. 2009. Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. Journal of the American Chemical Society 131, 6: 2122-2124.

Leikoski N, Fewer DP, Jokela J, Alakoski P, Wahlsten M & Sivonen K. 2012. Analysis of an inactive cyanobactin biosynthetic gene cluster leads to discovery of new natural products from strains of the genus *Microcystis*. PloS One 7, 8: e43002.

Leikoski N, Fewer DP, Jokela J, Wahlsten M, Rouhiainen L & Sivonen K. 2010. Highly diverse cyanobactins in strains of the genus *Anabaena*. Applied and Environmental Microbiology 76, 3: 701-709.

Leikoski N, Fewer DP & Sivonen K. 2009. Widespread occurrence and lateral transfer of the cyanobactin biosynthesis gene cluster in cyanobacteria. Applied and Environmental Microbiology 75, 3: 853-857.

Leikoski N, Liu L, Jokela J, Wahlsten M, Gugger M, Calteau A, Permi P, Kerfeld CA, Sivonen K & Fewer DP. 2013. Genome mining expands the chemical diversity of the cyanobactin family to include highly modified linear peptides. Chemistry & Biology 20, 8: 1033-1043.

Li Y-M, Milne JC, Madison LL, Kolter R & Walsh CT. 1996. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: Microcin B17 synthase. Science 274, 5290: 1188-1193.

Lindquist N, Fenical W, Van Duyne GD & Clardy J. 1991. Isolation and structure determination of diazonamides A and B, unusual cytotoxic metabolites from the marine ascidian *Diazona chinensis*. Journal of the American Chemical Society 113, 6: 2303–2304.

Long PF, Dunlap WC, Battershill CN & Jaspars M. 2005. Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production. ChemBioChem 6, 10: 1760-1765.

Matsuo Y, Kanoh K, Yamori T, Kasai H, Katsuta A, Adachi K, Shin-Ya K & Shizuri Y. 2007. Urukthapelstatin A, a novel cytotoxic substance from marine-derived *Mechercharimyces asporophorigenens* YM11-542. Journal of Antibiotics 60, 4: 251-260.

Mattila A, Andsten R-M, Jumppanen M, Assante M, Jokela J, Wahlsten M, Mikula KM, Sigindere C, Kwak DH, Gugger M, Koskela H, Sivonen K, Liu X, Yli-Kauhaluoma J, Iwaï H & Fewer D. 2019. Biosynthesis of the bis-prenylated alkaloids muscoride A and B. American Chemical Society Chemical Biology 14, 12: 2683-2690.

McIntosh JA, Donia MS, Nair SK & Schmidt EW. 2011. Enzymatic basis of ribosomal peptide prenylation in cyanobacteria. Journal of the American Chemical Society 133, 34: 13698-13705.

Mcintosh JA, Donia MS & Schmidt EW. 2010a. Insights into heterocyclization from two highly similar enzymes. Journal of the American Chemical Society 132, 12: 4089-4091.

McIntosh JA, Robertson CR, Agarwal V, Nair SK, Bulaj GW & Schmidt EW. 2010b. Circular logic: Nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst. Journal of the American Chemical Society 132, 44: 15499-15501.

McIntosh JA & Schmidt EW. 2010. Marine molecular machines: Heterocyclization in cyanobactin biosynthesis. ChemBioChem 11, 10: 1413-1421.

Mhlongo JT, Brasil E, de la Torre BG & Albericio F. 2020. Naturally occurring oxazole-containing peptides. Marine Drugs 18, 4: 203.

Molohon KJ, Melby JO, Lee J, Evans BS, Dunbar KL, Bumpus SB, Kelleher NL & Mitchell DA. 2011. Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics. American Chemical Society Chemical Biology 6, 12: 1307–1313.

Morita M, Hao Y, Jokela JK, Sardar D, Lin Z, Sivonen K, Nair SK & Schmidt EW. 2018. Post-translational tyrosine geranylation in cyanobactin biosynthesis. Journal of the American Chemical Society 140, 19: 6044-6048.

Muir JC, Pattenden G & Thomas RM. 1998. Total synthesis of (-)-muscoride A: A novel bis-oxazole based alkaloid from the cyanobacterium *Nostoc muscorum*. Synthesis, SPEC. ISS. APR.: 613-618.

Nagatsu A, Kajitani H & Sakakibara J. 1995. Muscoride A: A new oxazole peptide alkaloid from freshwater cyanobacterium *Nostoc muscorum*. Tetrahedron Letters, 36, 23: 4097-4100.

Newman DJ & Cragg GM. 2020. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. Journal of Natural Products 83, 3, pp. 770-803.

Paerl HW, Fulton 3rd. RS, Moisander PH & Dyble J. 2001. Harmful freshwater algal blooms, with an emphasis on cyanobacteria. The Scientific World Journal 1: 76-113.

Parajuli A, Kwak DH, Dalponte L, Leikoski N, Galica T, Umeobika U, Trembleau L, Bent A, Sivonen K, Wahlsten M, Wang H, Rizzi E, De Bellis G, Naismith J, Jaspars M, Liu X, Houssen W & Fewer DP. 2016. A unique tryptophan C-prenyltransferase from the Kawaguchipeptin biosynthetic pathway. Angewandte Chemie - International Edition 55, 11: 3596-3599.

Pettit RK. 2009. Mixed fermentation for natural product drug discovery. Applied Microbiology and Biotechnology 83, pp. 19-25.

Rippka R, Deruelles J, Waterbury JB, Herdman M & Stanier RY. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. Journal of General Microbiology 111, 1: 1-61

Sardar D, Hao Y, Lin Z, Morita M, Nair SK & Schmidt EW. 2017. Enzymatic N- and C-protection in cyanobactin RiPP natural products. Journal of the American Chemical Society 139: 2884−2887.

Sardar D, Pierce E, McIntosh JA & Schmidt EW. 2015. Recognition sequences and substrate evolution in cyanobactin biosynthesis. American Chemical Society Synthetic Biology 4, 2: 167-176.

Sardar D & Schmidt EW. 2016. Combinatorial biosynthesis of RiPPs: Docking with marine life. Current Opinion in Chemical Biology 31: 15-21.

Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG & Ravel J. 2005. Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. Proceedings of the National Academy of Sciences of the United States of America 102, 20: 7315-7320.

Schopf JW. 2012. The fossil record of cyanobacteria. In: Whitton BA (ed.). Ecology of cyanobacteria II: Their diversity in space and time. Netherlands: Springer. pp. 15–36.

Shin-ya K, Wierzba K, Matsuo K, Ohtani T, Yamada Y, Furihata K, Hayakawa Y & Seto H. 2001. Telomestatin, a novel telomerase inhibitor from *Streptomyces anulatus* [17]. Journal of the American Chemical Society 123, 6, pp. 1262−1263.

Singh R, Kumar M, Mittal A & Mehta PK. 2017. Microbial metabolites in nutrition, healthcare and agriculture. 3 Biotech 7, 1.

Sings HL & Rinehart KL. 1996. Compounds produced from potential tunicate-blue-green algal symbiosis: A review. Journal of Industrial Microbiology and Biotechnology 17, 5-6: 385-396.

Sivonen K. 2009. Cyanobacterial toxins. In: Moselio Schaechter (ed.). Encyclopedia of microbiology. Oxford: Elsevier. 3, pp. 290-307.

Sivonen K, Leikoski N, Fewer DP & Jokela J. 2010. Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. Applied Microbiology and Biotechnology 86, 5: 1213-1225.

Sohda K, Hiramoto M, Suzumura K, Takebayashi Y, Suzuki K & Tanaka A. 2005. YM-216391, a novel cytotoxic cyclic peptide from *Streptomyces nobilis*: II. Physico-chemical properties and structure elucidation. Journal of Antibiotics 58, 1: 32-36.

Suurnäkki S, Gomez-Saez GV, Rantala-Ylinen A, Jokela J, Fewer DP & Sivonen K. 2015. Identification of geosmin and 2-methylisoborneol in cyanobacteria and molecular detection methods for the producers of these compounds. Water research 68: 56-66.

Tan LT. 2013. Pharmaceutical agents from filamentous marine cyanobacteria. Drug Discovery Today 18, 17: 863-871.

Te SH, Tan FB, Boo CY, Thompson JR & Yew-Hoong Gin K. 2017. Genomics insights into production of 2-methylisoborneol and a putative cyanobactin by *Planktothricoides* sp. SR001. Standards in Genomic Sciences 12, 1: 35.

Yeh VSC. 2004. Recent advances in the total syntheses of oxazole-containing natural products. Tetrahedron 60, 52: 11995-12042.

Yorgey P, Lee J, Kördel J, Vivas E, Warner P, Jebaratnam D & Kolter R. 1994. Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor. Proceedings of the National Academy of Sciences of the United States of America 91, 10: 4519–4523.

Zhang Q-W, Lin L-G & Ye W-C. 2018. Techniques for extraction and isolation of natural products: a comprehensive review. Chinese Medicine 13, 20.

Ziemert N, Ishida K, Quillardet P, Bouchier C, Hertweck C, Tandeau de Marsac N & Dittmann E. 2008. Microcyclamide biosynthesis in two strains of *Microcystis aeruginosa*: from structure to genes and vice versa. Applied and Environmental Microbiology 74, 6: 1791–1797.

Velásquez JE & Van der Donk WA. 2011. Genome mining for ribosomally synthesized natural products. Current Opinion in Chemical Biology 15, 1: 11-21.

Walsby AE. 1994. Gas vesicles. Microbiological Reviews 58, 1: 94–144.

Welker M & Von Döhren H. 2006. Cyanobacterial peptides - Nature's own combinatorial biosynthesis. FEMS Microbiology Reviews 30, 4: 530-563.

Whitton BA & Potts M. 2012. Introduction to the cyanobacteria. In: Whitton BA (ed.) Ecology of cyanobacteria II: Their diversity in space and time. Netherlands: Springer. pp. 1–13.

Wipf P & Venkatraman S. 1996. Total synthesis of (-)-muscoride A. Journal of Organic Chemistry 61, 19: 6517-6522.

APPENDICES

Appendix 1: The characteristics for the nuclear magnetic resonance analysis assignments

Table S1. The nuclear magnetic resonance analysis assignments and their characteristics

Assignments*	Mixing time	Scans	Increments	Experiment time
^{1}H NMR	-	128	-	42 min
HMBC	-	64	360	15 h 56 min
COSY-45	-	8	400	2 h 8 min
TOCSY	100 ms	8	400	2 h 17 min
EASY ROESY	200 ms	32	400	9 h 10 min

*The proton chemical shift calibration was performed using the residual low-field pyridine signal (δH 8.74) as the reference and carbon chemical shift calibration was performed referencing Harris et al. (2002).

References

Harris RK, Becker ED, Cabral De Menezes SM, Goodfellow R & Granger P. 2002. NMR nomenclature: Nuclear spin properties and conventions for chemical shifts. IUPAC recommendations 2001. Solid State Nuclear Magnetic Resonance 22, 458–483.

Appendix 2: The solid phase extraction fractions from muscoride B purification

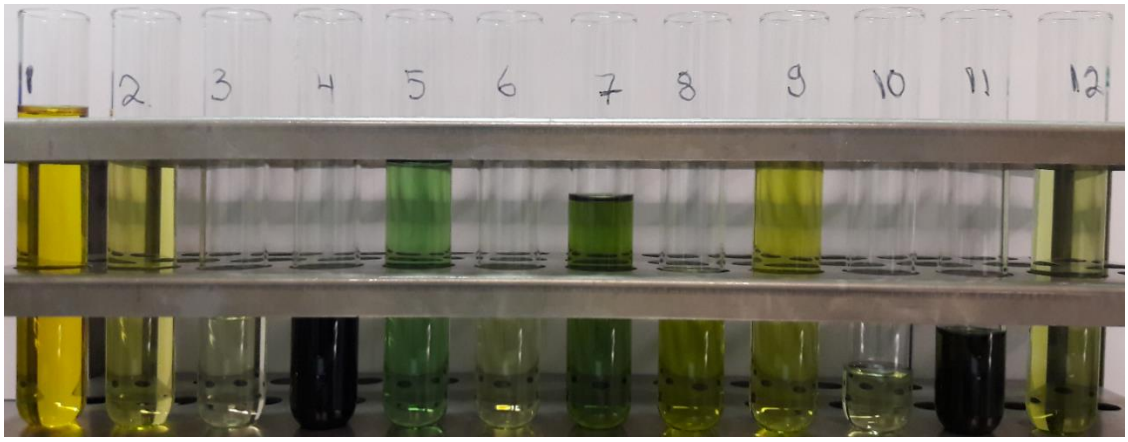


Figure S1. The SPE fractions from the first batch of lysed *Nostoc* sp. UHCC 0398 cells. Muscoride B eluted with methanol in fraction 11 and 12.

Figure S2. The SPE fractions from the second batch of lysed *Nostoc* sp. UHCC 0398 cells. Muscoride B eluted with methanol in fractions 8, 9 and 10.

Appendix 3: Nuclear magnetic resonance spectra of muscoride B

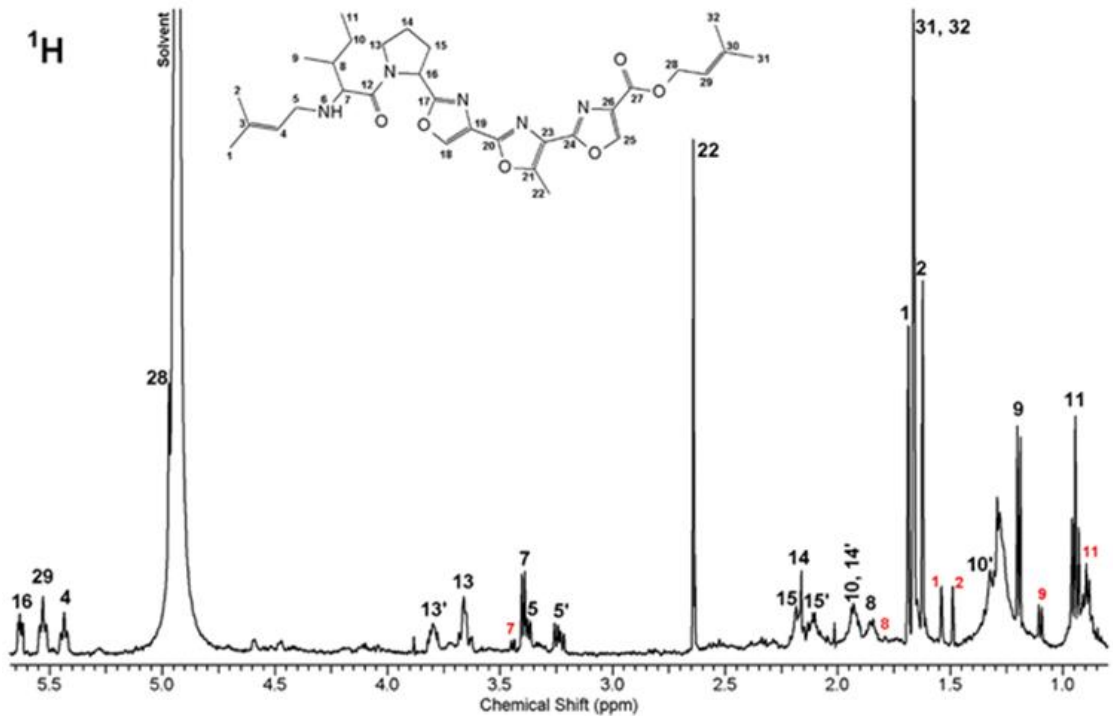


Figure S1. ^{1}H NMR spectrum of muscoride B (provided courtesy of Dr. Jouni Jokela).

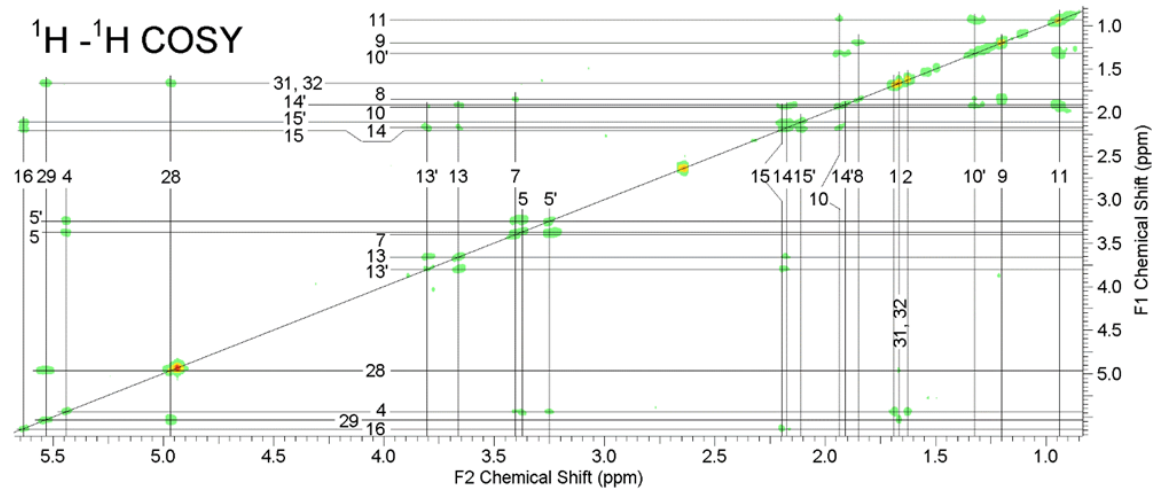


Figure S2. ^{1}H-^{1}H COSY spectrum of muscoride B (provided courtesy of Dr. Jouni Jokela).

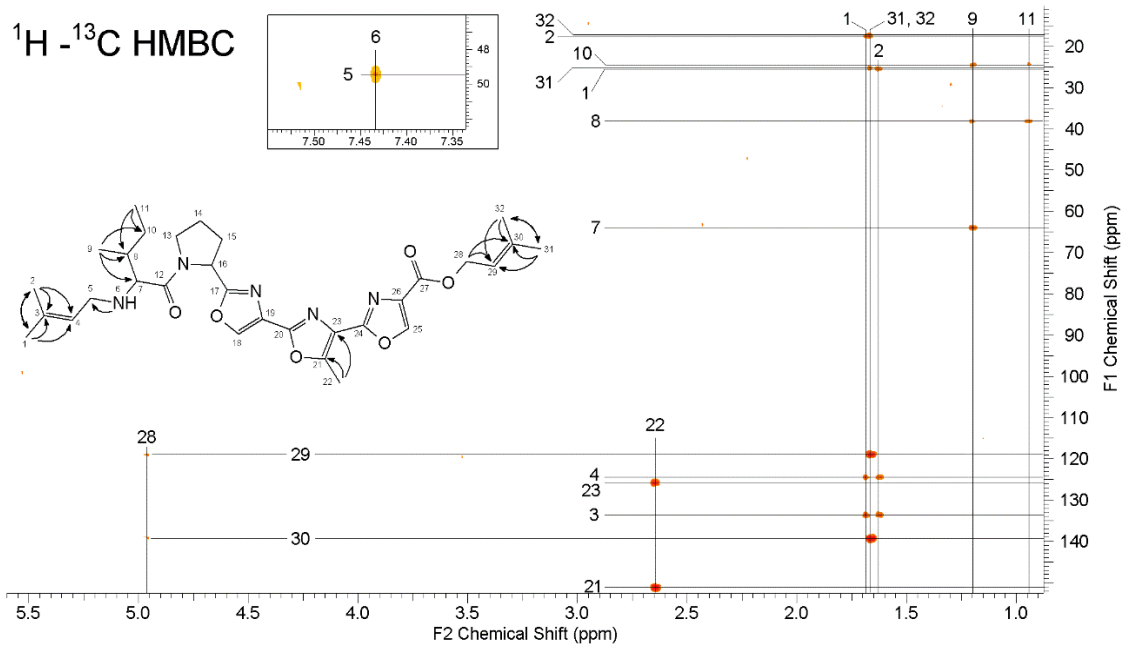


Figure S3. ^{1}H - ^{13}C HMBC spectrum of muscoride B (provided courtesy of Dr. Jouni Jokela).

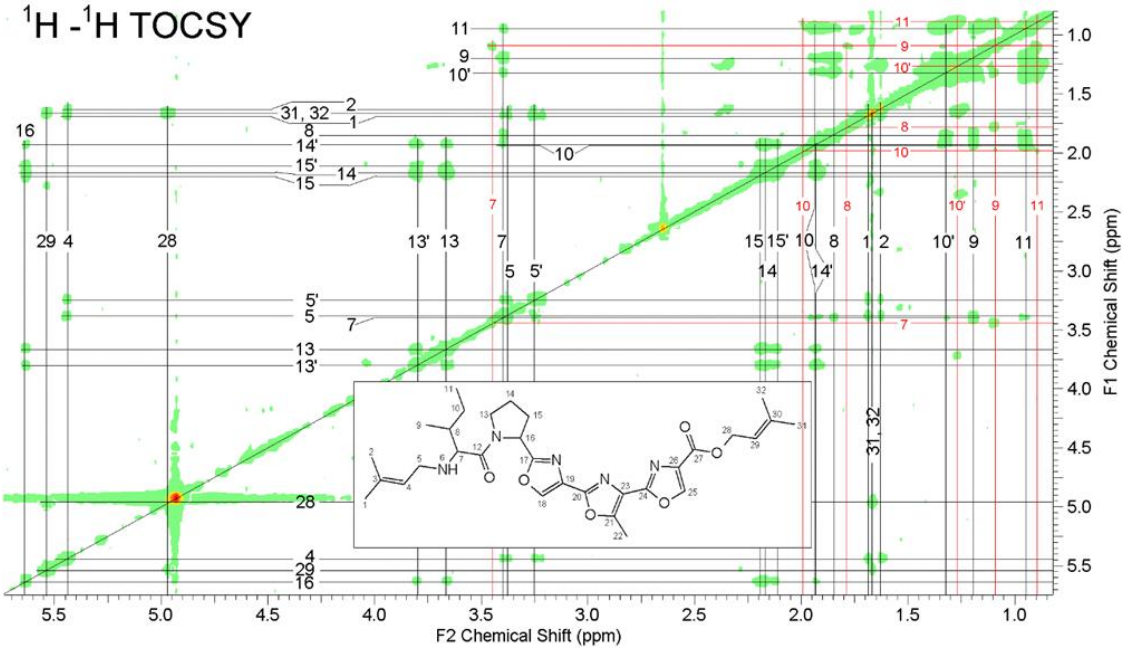


Figure S4. ^{1}H-^{1}H TOCSY spectrum of muscoride B (provided courtesy of Dr. Jouni Jokela).

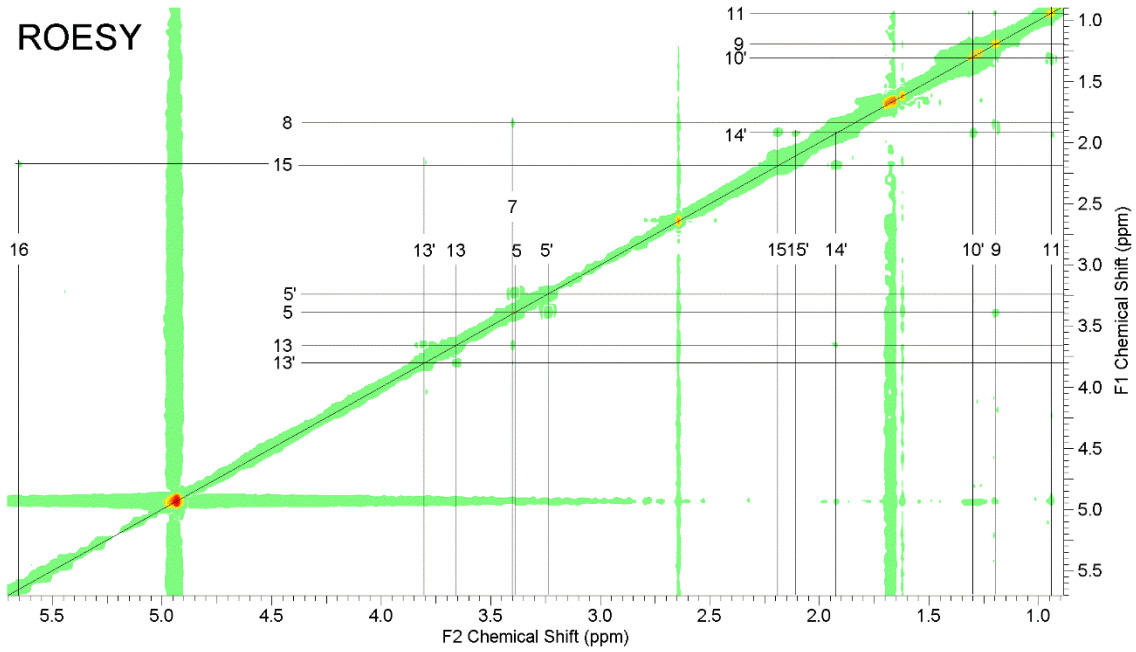


Figure S5. ROESY spectrum of muscoride B (provided courtesy of Dr. Jouni Jokela).

Appendix 4: Nuclear magnetic resonance spectral data for muscoride B

Table S1. ^{1}H and ^{13}C NMR spectral data for muscoride B major rotamer (80:20) in pyridine-d_5. Proton shifts for minor rotamer in parenthesis.

C-atom no	δ_C (ppm)	δ_H (ppm)	COSY[a]	TOCSY	ROESY	HMBC
Prenyl-1						
1	25.6	1.68 (1.54)	-	4, 5, 5’		2, 3, 4
2	17.9	1.62 (1.49)	-	4, 5, 5’		1, 3, 4
3	133.6	-	-	-	-	-
4	124.6	5.43	5, 5’	1, 2, 5, 5’		
5	49.4	3.38	4, 5’	1, 2, 4, 5’, 7	5’	
5’		3.24	4, 5	1, 2, 4, 5	5	
Ile						
6 (NH)	-	7.42	-	-		5
7	64.1	3.40 (3.44)	8	1, 4, 5’, 8, 9, 10, 10’, 11	8, 13	
8	38.3	1.85 (1.79)	9	7, 9, 10’, 11		
9	nd	1.20 (1.10)	8	7, 8, 10, 10’, 11	5, 8, 11	7, 8, 10
10	24.7	1.93 (1.98)	10’, 11	7, 9, 10’, 11		
10’		1.32 (1.26)	8, 10, 11	7, 8, 9, 10, 11	11, 14’	
11	nd	0.94 (0.90)	10, 10’	7, 8, 9, 10, 10’	10’, 14’	8, 10
12	nd	-	-	-	-	-
Pro						
13	nd	3.66	13’, 14, 14’	13, 14, 14’, 15, 15’, 16	13’	
13’		3.80	13, 14, 14’	13, 14, 14’, 15, 15’, 16	13, 15	
14	nd	2.17	13, 13’, 14’, 15’	13, 13’, 14, 15’, 16		
14’		1.92	14	13, 13’, 15, 15’, 16	13, 15	
15	nd	2.20	(13’), 14’, 15’, 16	13, 13’, 14’, 16	14’	
15’		2.11	14, 15	13, 13’, 14’, 15, 16	14’	
16	nd	5.63	15, 15’	13, 13’, 14, 14’, 15, 15’	15	
Tzl-1						
17	nd	-	-	-	-	-
18	nd	8.91*	-	-	-	-

19	nd	-	-	-	-	-
Tzl-2						
20	nd	-	-	-	-	-
21	151.1	-	-	-	-	-
22	11.2	2.64	-	-		21, 23
23	126.0	-	-	-	-	-
Tzl-3						
24	nd	-	-	-	-	-
25	nd	8.91*	-	-	-	-
26	nd	-	-	-	-	-
27	nd	-	-	-	-	-
Prenyl-2						
28	nd	4.96	29, (31, 32)	29, 31, 32		29, 30
29	119.0	5.53	28, (31, 32)	28, 31, 32		
30	139.3	-	-	-	-	-
31	25.3	1.66	(28, 29)	28, 29		29, 30, 32
32	17.7	1.67	(28, 29)	28, 29		29, 30, 31

a = couplings four/five bond away in parenthesis, nd = not detected, *= signal can be either of these atoms

Appendix 5: Peak areas for the prenylated products produced in the prenylation assay

Table S1. Integrated peak areas for the prenylated amino and carboxyl products from the prenylation assay analyzed with UPLC-HDMS

Prenylated product	Temperature (°C)	Reaction 1.	Reaction 2.
Amino			
Major (477)	28	179235	347147
Major (477)	37	150740	342306
Minor (140)	28	33,0	99
Minor (140)	37	31	104
Carboxy			
	28	5763	18514
	37	4363	14947